

**This Page Is Inserted by IFW Operations
and is not a part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- **BLACK BORDERS**
- **TEXT CUT OFF AT TOP, BOTTOM OR SIDES**
- **FADED TEXT**
- **ILLEGIBLE TEXT**
- **SKEWED/SLANTED IMAGES**
- **COLORED PHOTOS**
- **BLACK OR VERY BLACK AND WHITE DARK PHOTOS**
- **GRAY SCALE DOCUMENTS**

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2004, 16:23:59 ; Search time 48 Seconds
(without alignments)
3121.623 Million cell updates/sec

Title: US-09-964-238-2
Perfect score: 4991
Sequence: 1 MLTQAGNRKFKTECGKAF.....DGKMKTSKDHEDNEDMG 944

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

1:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4991	100.0	944	20	AAW83318
2	2612	52.3	533	22	ABW21853
3	2612	52.3	533	22	AAW57277
4	2612	52.3	533	22	AAW69682
5	2612	52.3	533	22	AAW30015
6	2612	52.3	533	22	ABG3298
7	1657.5	33.2	1125	24	ABU57636
8	1005	20.1	190	22	ABG48956
9	1005	20.1	190	22	ABG51171

10	1005	20.1	190	22	ABW28955	Peptide #1606 enco
11	1005	20.1	190	22	ABW31123	Peptide #3785 enco
12	1005	20.1	190	22	ABW34123	Peptide #1629 enco
13	1005	20.1	190	22	ABW36333	Peptide #3839 enco
14	1005	20.1	190	22	ABW19564	Protein #1563 enco
15	1005	20.1	190	22	ABW21680	Protein #3689 enco
16	1005	20.1	190	22	AAW4913	Human brain expres
17	1005	20.1	190	22	AAW57099	Human brain expres
18	1005	20.1	190	22	AAW67294	Human bone marrow
19	1005	20.1	190	22	AAW69491	Human bone marrow
20	1005	20.1	190	22	AAW15131	Peptide #1565 enco
21	1005	20.1	190	22	AAW17330	Peptide #3764 enco
22	1005	20.1	190	22	AAW27586	Peptide #1623 enco
23	1005	20.1	190	22	AAW29830	Peptide #3867 enco
24	1005	20.1	190	22	AAW02872	Peptide #1554 enco
25	1005	20.1	190	22	AAW05013	Peptide #3695 enco
26	1005	20.1	190	23	ABG36941	Human peptide enco
27	1005	20.1	190	23	ABG39120	Human peptide enco
28	817.5	16.4	747	22	ABW67181	Human peptide enco
29	817.5	16.4	1054	22	ABW57993	Drosophila melanog
30	717	14.4	141	19	AAW58369	Human CN238.1 prot
31	713	14.3	141	19	AAW58369	Human secreted pro
32	380.5	7.6	1252	22	AAW79739	Human protein SEO
33	379	7.6	1472	22	AAW1578	Novel human secret
34	374	7.5	1230	22	AAW30831	Novel human secret
35	371.5	7.4	1212	22	ABG00399	Novel human diagno
36	357	7.2	1342	22	ABG01726	Novel human diagno
37	356.5	7.1	1042	23	ABG96316	Human polypeptide
38	356.5	7.1	1395	23	ABG95054	Novel human diagno
39	355.5	7.1	1051	23	ABG96314	Novel human diagno
40	354	7.1	878	22	AAU29528	Human ovarian canc
41	350	7.0	1366	22	ABW59068	Novel human secret
42	349.5	7.0	923	22	ABW69037	Novel human diagno
43	349.5	7.0	1520	23	ABG27130	Human polypeptide
44	346.5	6.9	973	22	ABG21022	Novel human diagno
45	341.5	6.8	1939	23	ABW62852	Human polypeptide

ALIGNMENTS

RESULT 1	
AAW83318	
ID	AAW83318 standard; Protein, 944 AA.
XX	
AC	AAW83318;
XX	
XX	20-MAR-2003 (updated)
DT	01-MAR-1999 (first entry)
DE	Mouse SMAD interacting protein S1P1.
XX	
KW	S1P1; SMAD interacting protein 1; mouse; embryogenesis;
KW	differential; cancer; neural disease; immune disease; therapy;
KW	diagnosis.
OS	Mus sp.
XX	
FT	Key
FT	Region
XX	
XX	Location/Qualifiers
XX	166..216
XX	/note="Smad binding region"
XX	
XX	MO9855512-A2.
XX	
PD	10-DEC-1998.
XX	
XX	28-MAY-1998;
PF	98WO-EP03193.
XX	
XX	02-JUN-1997;
PR	97EP-0201645.
XX	
XX	(VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
PA	
XX	Huylebrock D, Remacle J, Verschueren K;
PI	

XX WPI; 1999-045732/04.
DR N-PSDB; AAV72502.

XX New SMAD interacting proteins obtained by a two hybrid screening
PT assay - useful for diagnosing developmental disorders and e.g.
PT cancer, immune and neural disease

XX Claim 7; Page 32-34; 48pp; English.

CC This is the amino acid sequence of a novel murine SMAD interacting
CC protein, designated S1P1. It is deduced from the sequence of a
CC cDNA clone (see AAV72502) isolated from mouse embryo cDNA using a
CC yeast two-hybrid screen. S1P1 represents a novel member of a
CC family of DNA-binding proteins including vertebrate delta-crystallin
CC enhancer binding proteins and Drosophila zfh-1. The C-terminal
CC zinc finger cluster of S1P1 (S1P1zf) binds to E2 box sites and to
CC the Brachyury protein binding site (see AAV72511), and interferes
CC with Brachyury-mediated transcription activation in cells. It also
CC interacts with the C-domain of Smad 1, 2 and 5. It may be involved
CC in transcriptional regulation of important differentiation genes in
CC significant biological processes such as cell growth and
CC differentiation, embryogenesis, and abnormal cell growth, including
CC cancer. The invention also provides recombinant expression vectors,
CC transfects or transduced cells, a method of screening for compounds
CC which affect the interaction between SMAD and SMAD interacting
CC protein, transgenic animals useful for testing medicaments and as
CC therapy models, and a method for post-transcriptional regulation of
CC gene expression by members of the TGF-beta superfamily by
CC manipulation or modulation of the interaction between Smad function
CC and/or activity and mRNA stability. S1Ps and nucleic acids encoding
CC them are useful therapeutically and in claimed kits for diagnosing
CC diseases or disorders related to cancer, malformation, immune or
CC neural disease, bone metabolism related diseases or disorders and
CC diseases affecting organs such as skin, lung, kidney, pancreas,
CC stomach, gonad, muscle or intestine.
CC (Updated on 20-MAR-2003 to correct PA field.)

CC Sequence 944 AA;

Query Match 100.0%; Score 4991; DB 20; Length 944;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 944; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTOGAGNKKFKTECGAKFYKHLKHLRHSGEKPECPNKKRFSHSGSYSHISS 60
DB 1 MTOGAGNKKFKTECGAKFYKHLKHLRHSGEKPECPNKKRFSHSGSYSHISS 60
QY 61 KKCIGLISVNGRRNNIKTGSSPNVSVPSTNSAITQLRNKLENGKPLSMSEQTGLIKIX 120
DB 61 KKCIGLISVNGRRNNIKTGSSPNVSVPSTNSAITQLRNKLENGKPLSMSEQTGLIKIX 120
QY 121 TEPILDNDYKVLMTATGFGSSPFNNNGIGATSPILGVHPSAOSPMOHLGVMGAPLILGPP 180
DB 121 TEPILDNDYKVLMTATGFGSSPFNNNGIGATSPILGVHPSAOSPMOHLGVMGAPLILGPP 180
QY 181 TMSNLSSEYQKVLQIDNTVSRQKMDCKTEDISKLYMHKDCSQPBEQGVTSPIPIPV 240
DB 181 TMSNLSSEYQKVLQIDNTVSRQKMDCKTEDISKLYMHKDCSQPBEQGVTSPIPIPV 240
QY 241 GLPVVSHNGATSIIDYTLKVNAAKCIQSLTTDSRRQISNIKKERLTLIDLVTDDKK 300
DB 241 GLPVVSHNGATSIIDYTLKVNAAKCIQSLTTDSRRQISNIKKERLTLIDLVTDDKK 300
QY 301 IENHSTSTPSCQFCESPPGPPILOHERYLCNNBEIKAVIQPHENIVPNAGVFN 360
DB 301 IENHSTSTPSCQFCESPPGPPILOHERYLCNNBEIKAVIQPHENIVPNAGVFN 360
QY 361 KALLISSVSEKGLTPIPIPYKDHMSVLYKAYYAMNPNPSDELKTSIANGVQBEVYKEM 420
DB 361 KALLISSVSEKGLTPIPIPYKDHMSVLYKAYYAMNPNPSDELKTSIANGVQBEVYKEM 420
QY 421 FEQRKVVQYNSRSPSLERTSKPLAPNSNPTTKDSLPRSPVYKPMDSITSPSIAELHNSV 480

DB 421 FEQRKVVQYNSRSPSLERTSKPLAPNSNPTTKDSLPRSPVYKPMDSITSPSIAELHNSV 480
QY 481 TSCDPLRLTKSSHFNINIKAVDKLDHSRNTGPSPLNLSSSTSSKNSHSSSTPSPSSSEL 540
DB 481 TSCDPLRLTKSSHFNINIKAVDKLDHSRNTGPSPLNLSSSTSSKNSHSSSTPSPSSSEL 540
QY 541 QAEPLDLSLPKQMRBEKGIATKNTKATISINDHNSVSSSSSENDEPLNLFPIKKEFSN 600
DB 541 QAEPLDLSLPKQMRBEKGIATKNTKATISINDHNSVSSSSSENDEPLNLFPIKKEFSN 600
QY 601 SNLNDKSNPNVPGNMPFSAKPLTLPPOQSAFPPTPMPVQTSIPGLRPYGLDQMSF 660
DB 601 SNLNDKSNPNVPGNMPFSAKPLTLPPOQSAFPPTPMPVQTSIPGLRPYGLDQMSF 660
QY 661 LPHMAVTPYTGATFADMOORRYOROGFOGDLDAQADYNSGLDMDTDSCLSRKXI 720
DB 661 LPHMAVTPYTGATFADMOORRYOROGFOGDLDAQADYNSGLDMDTDSCLSRKXI 720
QY 721 KTTESGMVACDLCDKTFQKSSSLRHKYEHNTGRPHQCQICKKAFGHKHLIHSRLHSG 780
DB 721 KTTESGMVACDLCDKTFQKSSSLRHKYEHNTGRPHQCQICKKAFGHKHLIHSRLHSG 780
QY 781 EKPYQCDKCGKRFSGSGYSQHMNRYSYCKBAEREAAREARKKHLGPTLLMNTA 840
DB 781 EKPYQCDKCGKRFSGSGYSQHMNRYSYCKBAEREAAREARKKHLGPTLLMNTA 840
QY 841 YLOSTIPQGSYSEERESMPROGESKEHEKGESEGYGLRRDGESESESESESENS 900
DB 841 YLOSTIPQGSYSEERESMPROGESKEHEKGESEGYGLRRDGESESESESESENS 900
QY 901 MDTPETIRDEERTGDHSMDSSEDGKMETKSDHEEDNNEDGNG 944
DB 901 MDTPETIRDEERTGDHSMDSSEDGKMETKSDHEEDNNEDGNG 944

RESULT 2

ABB21853

ID ABB21853 standard; Protein; 533 AA.

XX ABB21853;

XX 23-JAN-2002 (first entry)

XX

Protein #3852 encoded by probe for measuring heart cell gene expression.

KW Human; gene expression; heart; microarray; vascular system;

KW cardiovascular disease; hypertension; cardiac arrhythmia;

XX Homo sapiens.

PN MO200157274-A2.

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WC-US00666.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488899/53.

XX

Single exon nucleic acid probes for analyzing gene expression in human

PT hearts -
XX
PS Claim 15; SEQ ID No 23623; 530bp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC AB21535-ABA41305). The present sequence is a protein encoded by one such
CC probe. The probe may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosticating diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPRO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 533 AA;
Query Match 52.3%; Score 2612; DB 22; Length 533;
Best Local Similarity 94.2%; Pred. No. 5.3e-177;
Matches 502; Conservative 12; Mismatches 19; Indels 0; Gaps 0;
QY 36 EKPYECNCKKRFHSHSGSYSSHISKKKICIGLISVNGMRNNIKTGSSPNSVSSPTNSAI 95
DB 1 EKPYECNCKKRFHSHSGSYSSHISKKKICIGLISVNGMRNNIKTGSSPNSVSSPTNSAI 60
QY 96 TOLRNKLENGKPLMSSEOTGLIKTEPLDPNDYKVLMAHGFSGSSPNNGGIGATSP 155
DB 61 TOLRNKLENGKPLMSSEOTGLIKTEPLDPNDYKVLMAHGFSGSSPNNGGIGATSP 120
QY 156 GVHPSAOSPMOHLGVGMEAPLGFPTMNSNLSEVOKVLOIVDNTVSRQMDCKTEDISK 215
DB 121 GVHPSAOSPMOHLGVGMEAPLGFPTMNSNLSEVOKVLOIVDNTVSRQMDCKTEDISK 180
QY 216 KGYHMKDPCSQPEBQGVTSPIPVGLPVVSHNGATKSIIDYTELEKNEAACAQSLTTD 275
DB 181 KGYHMKDPCSQPEBQGVTSPIPVGLPVVSHNGATKSIIDYTELEKNEAACAQSLTTD 240
QY 276 SRROIINIKKKKRLTLDLVTDKMIENHSTPSCQFCESPPGPIPLHOHERYLYCKM 335
DB 241 SRROIINIKKKKRLTLDLVTDKMIENHSTPSCQFCESPPGPIPLHOHERYLYCKM 300
QY 336 NEIKAVLQPHENIVPNKAGVFNKALLSSVSEKGLTSPINPYKDHMSVLAAYYAMN 395
DB 301 NEIKAVLQPHENIVPNKAGVFNKALLSSVSEKGLTSPINPYKDHMSVLAAYYAMN 360
QY 396 MEPSDELKISIAVGLPQEFVKEMFEQRKYQYNSRSRPSLERTSKPLAPNSPTTKDS 455
DB 361 MEPSDELKISIAVGLPQEFVKEMFEQRKYQYNSRSRPSLERTSKPLAPNSPTTKDS 420
QY 456 LLPRSPVKPMDSITSPSIAELHNSVTSCDPLRLTKSHFTNIKAVDKLHSRNTSP 515
DB 421 LLPRSPVKPMDSITSPSIAELHNSVTSCDPLRLTKSHFTNIKAVDKLHSRNTSP 480
QY 516 NLSTSSKNSHSSSYTPNSFSSEELQAEPLDLSPKORREPKGIATKNTKTA 568
DB 481 NLSTSSKNSHSSSYTPNSFSSEELQAEPLDLSPKORREPKGIATKNTKTA 533
RESULT 3
ID AAM57277 standard; Protein: 533 AA.
XX AAM57277;
AC AAM57277;
DT 05-NOV-2001 (first entry)
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 29382.
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 29382.
KW Human, brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.

XX Homo sapiens.
OS
XX
XX WO200157275-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00667.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-483446/52.
DR
XX
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
PT
XX
XX Example 4; SEQ ID NO: 29382; 650bp + Sequence Listing; English.
PS
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
SQ Sequence 533 AA;
Query Match 52.3%; Score 2612; DB 22; Length 533;
Best Local Similarity 94.2%; Pred. No. 5.3e-177;
Matches 502; Conservative 12; Mismatches 19; Indels 0; Gaps 0;
QY 36 EKPYECNCKKRFHSHSGSYSSHISKKKICIGLISVNGMRNNIKTGSSPNSVSSPTNSAI 95
DB 1 EKPYECNCKKRFHSHSGSYSSHISKKKICIGLISVNGMRNNIKTGSSPNSVSSPTNSAI 60
QY 96 TOLRNKLENGKPLMSSEOTGLIKTEPLDPNDYKVLMAHGFSGSSPNNGGIGATSP 155
DB 61 TOLRNKLENGKPLMSSEOTGLIKTEPLDPNDYKVLMAHGFSGSSPNNGGIGATSP 120
QY 156 GVHPSAOSPMOHLGVGMEAPLGFPTMNSNLSEVOKVLOIVDNTVSRQMDCKTEDISK 215
DB 121 GVHPSAOSPMOHLGVGMEAPLGFPTMNSNLSEVOKVLOIVDNTVSRQMDCKTEDISK 180
QY 216 KGYHMKDPCSQPEBQGVTSPIPVGLPVVSHNGATKSIIDYTELEKNEAACAQSLTTD 275
DB 181 KGYHMKDPCSQPEBQGVTSPIPVGLPVVSHNGATKSIIDYTELEKNEAACAQSLTTD 240
QY 276 SRROIINIKKKKRLTLDLVTDKMIENHSTPSCQFCESPPGPIPLHOHERYLYCKM 335
DB 241 SRROIINIKKKKRLTLDLVTDKMIENHSTPSCQFCESPPGPIPLHOHERYLYCKM 300
QY 336 NEIKAVLQPHENIVPNKAGVFNKALLSSVSEKGLTSPINPYKDHMSVLAAYYAMN 395
DB 301 NEIKAVLQPHENIVPNKAGVFNKALLSSVSEKGLTSPINPYKDHMSVLAAYYAMN 360
QY 396 MEPSDELKISIAVGLPQEFVKEMFEQRKYQYNSRSRPSLERTSKPLAPNSPTTKDS 455
DB 361 MEPSDELKISIAVGLPQEFVKEMFEQRKYQYNSRSRPSLERTSKPLAPNSPTTKDS 420
QY 456 LLPRSPVKPMDSITSPSIAELHNSVTSCDPLRLTKSHFTNIKAVDKLHSRNTSP 515
DB 421 LLPRSPVKPMDSITSPSIAELHNSVTSCDPLRLTKSHFTNIKAVDKLHSRNTSP 480

QY	516	ILSTSSKNSKNSSSSYTPNSFSSSELOAPDLISLPKORREKGIATNKTKA	568
Db	481	ILSTSSKNSKNSSSSYTPNSFSSSELOAPDLISLPKORREKGIATNKTKA	533
RESULT 4			
QY	AA669682	standard; Protein; 533 AA.	
Db	AA669682		
QY	06-NOV-2001	(first entry)	
Db	Human bone marrow expressed probe encoded protein SEQ ID NO: 29988.		
QY	Human; bone marrow expressed exon; gene expression analysis; probe;		
Db	microarray; cancer; leukaemia; lymphoma; myeloma.		
QY	Homo sapiens.		
Db	MO200157276-A2.		
QY	09-AUG-2001.		
Db	30-JAN-2001; 2001MO-US000668.		
QY	04-FEB-2000; 2000US-0180312.		
Db	26-MAY-2000; 2000US-0207456.		
QY	30-JUN-2000; 2000US-0608408.		
Db	03-AUG-2000; 2000US-0632366.		
QY	21-SEP-2000; 2000US-0234687.		
Db	27-SEP-2000; 2000US-0236359.		
QY	04-OCT-2000; 2000GB-0024263.		
Db	(MOLE-) MOLECULAR DYNAMICS INC.		
QY	Penn SG, Hanzel DK, Chen W, Rank DR;		
Db	WPI; 2001-488990/53.		
QY	Human genome-derived single exon nucleic acid probes useful for		
Db	analyzing gene expression in human bone marrow -		
QY	Example 4; SEQ ID NO: 29988; 658bp + Sequence Listing; English.		
Db	The present invention provides a number of single exon nucleic acid		
QY	probes which are derived from genomic sequences expressed in the human		
Db	bone marrow. They can be used to measure gene expression in bone marrow		
QY	samples, which may enable the improved diagnosis and treatment of cancers		
Db	such as lymphoma, leukaemia and myeloma. The present sequence is a		
QY	protein encoded by one of the probes of the invention.		
Db	Sequence 533 AA.		
QY	Query Match 52.3%; Score 2612; DB 22; Length 533;		
Db	Best Local Similarity 94.2%; Pred. No. 5.3e-177;		
QY	Matches 502; Conservative 12; Mismatches 19; Indels 0; Gaps 0;		
Db	1 EKPECPNCKRRFSHSGSYSHISKCKIGLISVNGRRNNKITSSSPNSVSSSPTNSAI	95	
QY	TOLRNKLNGKPLMSSEOTGLLKITEPLDFNDYKVLMAATHGFGSSSPMNGLGATSP	155	
Db	61 TOLRNKLNGKPLMSSEOTGLLKITEPLDFNDYKVLMAATHGFGSGTSPMNGLGATSP	120	
QY	156 GVHPAAGSPMHLGVMEAPLLGFPPTANSNISEVQKVIQIVDNTVSRQMDCKEDISK	215	
Db	121 GVHPAAGSPMHLGVMEAPLLGFPPTANSNISEVQKVIQIVDNTVSRQMDCKAEISK	180	
QY	KGIYHMDCSQPEEGVTSNIPPIPLGVVSNNGATKSIIDTLEKVNFAKCIOSLT	275	

Dd		18	KGYHMKDPCSGDEQGVTSPINPVGLPVSHNGATSKSIIDYLEKYNKAACIQSLTTD	246
Oy		276	SRRQISNIKEKRLRTLDLVYDDKKIENHSISTPSCFCESPPGIPLHHORYLCRM	3353
Dd		241	SRROISNIEKERLRTLDLVYDDKKIENHNISTPSPCCFCESEPGPIPLOHORYLCRM	300
Oy		336	NERIKAVLDPHENIYPNKAGVPVDKALLSSVYSEKGLTSPINPYODHMSVLKAYYAMN	395
Dd		301	NBIIRKAVLPDENIYPNKAGVPVDKALLSSVYSEKGMTSPINPYODHMSVLKAYYAMN	366
Oy		396	MEBNSDELTKISIAVGLPDFEFVKWFEQRKYQYQSNRSPSLESTSKPLAPNSNPTKDS	455
Dd		361	MEDNSDELTKISIAVGLPDFEFVKWFEQRKYQYQSNRSPSELERSKPLAPNSNPTKDS	420
Oy		456	LILPRSVKRWMDISTPSIAELHNSVTCDPPLRLTKSHFNITKAVDKLDRSNTSPL	515
Dd		421	LILPRSVKRWMDISTPSIAELHNSVTCDPPLRLTKSPHFNIKRVKELDRSNTSPL	480
Oy		516	NLSSTSKSHSSSYTPNSFSSEEQAQLDLSLPKWRBEKGIIATKNKTKA	568
Dd		481	NLSSTSKSHSSSYTPNSFSSEEQAQLDLSPFLVIITKMERIQKYSHKEONKS	533
RESULT 5				
ID	AAM30015			
XX	AAM30015 standard; Protein; 533 AA.			
AC	AAM30015;			
AD				
DT	17-OCT-2001 (first entry)			
XX				
DE	Peptide #4052 encoded by probe for measuring placental gene expression.			
DM				
XX	Probe; microarray; human; placenta; antenatal diagnosis;			
KM	genetic disorder.			
XX				
OS	Homo sapiens.			
XX				
PN	WO200157272-A2.			
PD				
XX	09-AUG-2001.			
PF				
XX	30-JAN-2001; 2001MO-USO0663.			
XX				
PR	04-FEB-2000; 2000US-0180312.			
PR	26-MAY-2000; 2000US-0207456.			
PR	30-JUN-2000; 2000US-0608408.			
PR	03-AUG-2000; 2000US-0632366.			
PR	21-SEP-2000; 2000US-0234687.			
PR	27-SEP-2000; 2000US-0236359.			
PR	04-OCT-2000; 2000GB-0024263.			
XX				
PA	(MOL-E-) MOLECULAR DYNAMICS INC.			
XX				
XE	Penn SG, Hanzel DK, Chen W, Rank DR;			
XX				
DR	WPI; 2001-488897/53.			
XX				
PT	Human genome-derived single exon nucleic acid probes useful for			
PT	analyzing gene expression in human placenta -			
XX				
XX	Claim 27; SEQ ID No 30284; 654bp; English.			
CC	The present invention relates to single exon nucleic acid probes (SENPs:			
CC	see A113315-A157546). The present sequence is a peptide encoded by one			
CC	such probe. The probes are useful for producing a microarray for			
CC	predicting, measuring and displaying gene expression in samples derived			
CC	from human placenta. The probes are useful for antenatal diagnosis of			
CC	human genetic disorders.			
XX				
XX				
SO	Sequence 533 AA;			

Query Match 52.3%; Score 2612; DB 22; Length 533;

Best Local Similarity 94.2%; Pred. No. 5.3e-177;
Matches 502; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

```

QY 36 EKPYECPCCKRPFSSGSSYSSHISKKCTIGLISVNGRRNNIKTGSSPNSVSSPTNSAI 95
    |||
DB 1 EKPYECPCCKRPFSSGSSYSSHISKKCTIGLISVNGRRNNIKTGSSPNSVSSPTNSAI 60
QY 96 TOLRNKLENGKPLSMSEQTGLIKITEPLDFNDYKVLMTATGFGSSPPFMMGIGATSP 155
    |||
DB 61 TOLRNKLENGKPLSMSEQTGLIKITEPLDFNDYKVLMTATGFGSGTSPFMMGIGATSP 120
QY 156 GVHPSAOSPMDHGVGMEAPLIGPPTMNSNSEVOKLQIVDNTVSRQKMDCKEDISK 215
    |||
DB 121 GVHPSAOSPMDHGVGMEAPLIGPPTMNSNSEVOKLQIVDNTVSRQKMDCKEDISK 180
QY 216 KGYHMKDPCSQPEEGQVTSFNIIPVGLPVVSHNGATKSIIDYTLKVEAKACIQSLT 275
    |||
DB 181 KGYHMKDPCSQPEEGQVTSFNIIPVGLPVVSHNGATKSIIDYTLKVEAKACIQSLT 240
QY 276 SRRQISNIKKERLRLIDLVTDKMIENHSISTPSCQCKESPFGPIPLHQHERYLCKM 335
    |||
DB 241 SRRQISNIKKERLRLIDLVTDKMIENHSISTPSCQCKESPFGPIPLHQHERYLCKM 300
QY 336 NEETKAVIQPHENIVPNKAGVFDNKALLSSVLSSEKLTSPINPYKDHMSVLYKAYTAN 395
    |||
DB 301 NEETKAVIQPHENIVPNKAGVFDNKALLSSVLSSEKMTSPINPYKDHMSVLYKAYTAN 360
QY 396 MEPSDELKTSIANGLPQEPFKMEFEQRKAYOVSNSPSPISERTSKLANSNPPTKDS 455
    |||
DB 361 MEPSDELKTSIANGLPQEPFKMEFEQRKAYOVSNSPSPISERTSKLANSNPPTKDS 420
QY 456 LLPSRPVPMDSITSPSIAELHNSVTSQDPLRLTKSSHFTNIKAVDKLDSRSNTSP 515
    |||
DB 421 LLPSRPVPMDSITSPSIAELHNSVTSQDPLRLTKSSHFTNIKAVDKLDSRSNTSP 480
QY 516 NLSTSSKNSHSSSTTTPNSFSSEELQAPPLDLSLPKQREPKGIATNTKTKA 568
    |||
DB 481 NLSTSSKNSHSSSTTTPNSFSSEELQAPPLDLSLPKQREPKGIATNTKTKA 533

```

RESULT 6

ABG39298
ABG39298 standard; Peptide; 533 AA.

AC ABG39298;

DT 19-AUG-2002 (first entry)

DE Human peptide encoded by genome-derived single exon probe SEQ ID 28963.

XX Human: single exon probe; asthma; lung cancer; COPD; ILD;
KM chronic obstructive pulmonary disease; interstitial lung disease;
KM familial idiopathic pulmonary fibrosis; neurofibromatosis;
KM tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;
KM Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KM pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;
KM pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KM primary ciliary dyskinesia; pulmonary hypertension;
KM hyaline membrane disease.

XX Homo sapiens.

OS WO200186003-A2.

PD 15-NOV-2001.

XX 30-JAN-2001; 2001WO-US00665.

XX 04-FEB-2000; 2000US-180312P.

XX 26-MAY-2000; 2000US-207456P.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-234687P.

PR 27-SEP-2000; 2000US-236359P.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2002-114183/15.

PT Spatially-addressable set of single exon nucleic acid probes, used to

PS measure gene expression in human lung samples -

CC Claim 27; SEQ ID No 28963; 634pp; English.

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridize at high stringency to a nucleic acid expressed in the human lung; measuring gene expression in a sample derived from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarray having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberosus sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

SQ Sequence 533 AA;

Query Match 52.3%; Score 2612; DB 23; Length 533;

Best Local Similarity 94.2%; Pred. No. 5.3e-177;
Matches 502; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

```

QY 36 EKPYECPCCKRPFSSGSSYSSHISKKCTIGLISVNGRRNNIKTGSSPNSVSSPTNSAI 95
    |||
DB 1 EKPYECPCCKRPFSSGSSYSSHISKKCTIGLISVNGRRNNIKTGSSPNSVSSPTNSAI 60
QY 96 TOLRNKLENGKPLSMSEQTGLIKITEPLDFNDYKVLMTATGFGSSPPFMMGIGATSP 155
    |||
DB 61 TOLRNKLENGKPLSMSEQTGLIKITEPLDFNDYKVLMTATGFGSGTSPFMMGIGATSP 120
QY 156 GVHPSAOSPMDHGVGMEAPLIGPPTMNSNSEVOKLQIVDNTVSRQKMDCKEDISK 215
    |||
DB 121 GVHPSAOSPMDHGVGMEAPLIGPPTMNSNSEVOKLQIVDNTVSRQKMDCKEDISK 180

```

QY 216 KGYHMDPCSQPEQGVTSNPINPVGLPVVSHNGATKSIIDYTLKYNKAKCLOSLTID 275
 DB 181 KGYHMDPCSQPEQGVTSNPINPVGLPVVSHNGATKSIIDYTLKYNKAKCLOSLTID 240
 QY 276 SRROINIKKEKRTTIDLTVDKMIENHSISTPSCQFCKESPPGPIPLHGERLYCKM 335
 DB 241 SRROINIKKEKRTTIDLTVDKMIENHSISTPSCQFCKESPPGPIPLHGERLYCKM 300
 QY 336 NEEIKAVLOPHENIVPNKAGVFNKALLLSVLSSEKLTSPINPYKDHMSVLKAYYAMN 395
 DB 301 NEEIKAVLOPHENIVPNKAGVFNKALLLSVLSSEKLTSPINPYKDHMSVLKAYYAMN 360
 QY 396 MEENSELKISTAVGLPOEFVKEWEPQKRVQYNSRSPLERTSGPLAPNSNPTTKS 455
 DB 361 MEENSELKISTAVGLPOEFVKEWEPQKRVQYNSRSPLERTSGPLAPNSNPTTKS 420
 QY 456 LIPRSFVKPMDISTSPSIALHNSVTSQDPLRLTSSHFTNIKAVDKLDHSRNTSPL 515
 DB 421 LIPRSFVKPMDISTSPSIALHNSVTSQDPLRLTSSHFTNIKAVDKLDHSRNTSPL 480
 QY 516 NLSSTSSKNSHSSSYTPNSFSSEELQAEPLDLSLPKQMRPEKGI IATKNTKA 568
 DB 481 NLSSTSSKNSHSSSYTPNSFSSEELQAEPLDLSLPKQMRPEKGI IATKNTKA 533
 RESULT 7
 ID ABUS7636 standard; Protein; 1125 AA.
 AC ABUS7636;
 DT 09-APR-2003 (first entry)
 DE Differentially expressed breast cancer associated protein #23.
 XX Breast cancer; differential gene expression; BC-cDNA;
 KM breast cancer diagnosis; breast cancer monitoring;
 KM breast cancer treatment; breast cancer staging.
 XX Homo sapiens.
 OS US2002156263-A1.
 PN 24-OCT-2002.
 PD 04-OCT-2001; 2001US-0974298.
 PF 05-OCT-2000; 2000US-238331P.
 PR (CHEN/) CHEN H.
 PA Chen H;
 PI WPI; 2003-182653/18.
 DR New cDNA, which are differentially expressed in (metastatic) breast
 XX cancer useful for diagnosing or staging breast cancer, or for
 PT monitoring the treatment of breast cancer in an individual -
 PS Example; SEQ ID NO 114; 30pp; English.
 XX The invention describes a combination of cDNAs (designated BC-cDNAs),
 CC which are differentially expressed in breast cancer. The combination
 CC includes 152 cDNA sequences, or their complements. The protein encoded
 CC by any of these BC-cDNAs is useful for screening several molecules or
 CC compounds to identify at least one ligand that specifically binds the
 CC protein, producing or preparing polyclonal or monoclonal antibodies, or
 CC purifying antibodies from a sample. The antibodies, which specifically
 CC bind the protein differentially expressed in breast cancer is useful for
 CC detecting the expression of a protein in a sample. The BC-cDNAs are
 CC also useful for diagnosing, monitoring the treatment of, or staging,
 CC breast cancer. This is the amino acid sequence of a differentially
 CC expressed breast cancer associated protein.

CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from the US patent office at
 CC seqdata.uspto.gov/sequence.html?docid=20020156263.

SO Sequence 1125 AA;

Query Match 33.2%; Score 1657.5; DB 24; Length 1125;
 Best Local Similarity 41.9%; Pred. No. 1.2e-108;
 Matches 406; Conservative 147; Mismatches 308; Indels 107; Gaps 32;

QY 2 LFGAGNRKFKCTCEGAKFYKHLKHLRHSXGKPYCPNCKRFSHSGSYSHISK 61
 DB 232 VTQSGCNRKFKCTCEGAKFYKHLKHLRHSXGKPYCPNCKRFSHSGSYSHISK 291
 QY 62 KICGLISVNGRMRNNIKTG--SSPNSVSSPTNSAITQLRNKLNGKPLSMSEQTLKI 119
 DB 292 KICSLIPVNGRPRRTGLKTSQCSSP--SLSPSPSPRPQROKTEH--KPL--QEQLSVNGI 347
 QY 120 KTEPLDFNDYKVLMTATGFSGSSPPFANGGLATSPGLVHPSPASQSPQHL---GYGAEPL 176
 DB 348 KTEPLVDY--EFKPIVVASGINGCSTPLONGVFTGGGLQATSSPQGWQAVVLPVGLVSP 406
 QY 177 LGPPTNNSLSEYQKVLQI--VDNTVSRQKMDCKTEDISKYGYHMDPCSQPEQGVTS 235
 DB 407 -----SINLSDIQNVKAVDGNVIRQVLE--NNQALASKQETIASEPIQCGSHV 457
 QY 236 NIPVGLPVVSHNGATKSIIDYTLKYNKAKCLOSLTIDSRROINIKKEKRTTIDLV 295
 DB 458 -ISALSIPVLDQGTGKIITINYSLEQPSQLQVPPQLKKNPVATNSCKSEKLPEDLYVK 516
 QY 296 TD-DKMIENH-SISTPSCQFCKESPPGPI-PLHGERLYCKMNEIKAVLOPHENIVPN 352
 DB 517 SEKDKSPEGGVNDSTLLCDDC---PGDINALPELKHV-----DLKOPTP---BP 561
 QY 353 KAGVFNKALLLSVLSSEKLTSPINPYKDHMSVLKAYYAMMEENSELKISTAVGL 412
 DB 562 LPAAEKREBSYSSATGCGNLSPOPLKILSLIKAYALMAQSAEELSIALDSVNL 621
 QY 413 POEFVKEWEPQKRVQYNSRSPLERTSGPLAPNSNPTTKOSLIPRSFVKPMDISTSPS 472
 DB 622 PLDVVKWEPKQAGQIS-----VQSESPSPBERGV-----NTPAKNNDQQOSAN 667
 QY 473 IAEIHSVTSQDPLRLTSSHFTNIKAVDKLDHSRNTSPLNLSSTSSKNSHSSSYTP 532
 DB 668 ANEPQDSTVNLQSLPMKTNSS---PVLPGVSTTNGSSSTSPSPPLNLSSTRTQGLVYA 724
 QY 533 NSPFSSELOAEPLDLSLPKQMRPEKGI IATKNTKATSIINLHNSVSSSENS-----DE 587
 DB 725 EG-AQEEPOVEPLDLSLPKQGE-----LLERSTTIVYQNSVYSVOGE 767
 QY 588 PLNLTFIKKEFSNNSNLDNKNPVFGMNPFSAPLYTLPPOSAPFPATFMPVQTSIP 647
 DB 768 PLNLSCAKKEPQDSCVTD--SEPVNVNIPPSANPINIAPVTAQLPTIVAIADQNSVP 825
 QY 648 GLRPFYGLDQMSLPLMAATTPYGAATPFADMQQRKRYQKQGGQSLLDGADYMGGLD 707
 DB 826 CRLALANKQTLIPQVATYVTVSP-AVOEPLPLVIOPNGQDERQOTSBSGVSND 884
 QY 708 MTPDSCLSRKKIKKTESGVYACDLCTFKFQSSSLRHKYEHTEGRPHQCOCKKAFKH 767
 DB 885 QNDSDSSTPPKKAARKTENGMVACDLCDKIFQSSSLRHKYEHTEGRPHQCGCKKAFKH 944
 QY 768 KHLIHSRLHSXGKPYQDKCGKRFSGSYSGQHNNHRYSYCKREAEAREAREARK 827
 DB 945 KHLIHSRLHSXGKPYQDKCGKRFSGSYSGQHNNHRYSYCKREAEAREAREARK 1000
 QY 828 GHLPYGLMLNRAVLOSITP-QGYSSERESKPR--DSESEKHEKGE-----GYK 879
 DB 1001 -EAGP-EIISNETHVAGARAPSQG--DSDERESTITREDDSEKEEEDKEMELQOEKE 1056
 QY 880 LRRDDEDEEESEKSMQTDPEITRDEEETDHSMDSSSD-----GKMET 930

Db 1057 CEKPGDEEEEEEVEEVE-FAENGBEATKGLKODRAESQASSLGQKVESSE 1115

Qy 931 KSDHEEDN 938

Db 1116 QVSEKTN 1123

RESULT 8

ABG48956

ID ABG48956 standard; Peptide; 190 AA.

XX

AC ABG48956;

XX

DT 25-FEB-2003 (first entry)

XX

DE Human liver peptide, SEQ ID No 27604.

XX

KM Human, liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;

KM hypercholesterolaemia; coronary heart disease.

XX

OS Homo sapiens.

XX

PN WO200157273-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00664.

XX

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488898/53.

XX

PT Human genome-derived single exon nucleic acid probes useful for

XX analysing gene expression in human adult liver -

PS Claim 27; SEQ ID No 27604; 658bp; English.

XX

CC The invention relates to a single exon nucleic acid probe (SENP) (1) for

CC measuring human gene expression in a sample derived from human adult

CC liver, comprising one of 13109 defined nucleotide sequences given in the

CC specification (or complements/ fragments). The probe hybridises at high

CC stringency to a nucleic acid molecule expressed in the human adult

CC liver. (1) may be used for predicting, measuring and displaying gene

CC expression in samples derived from human adult liver. The genes

CC identified may be involved in genetic liver diseases such as cirrhosis,

CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which

CC is associated with coronary heart disease. ABG47348-ABG5930 represent

CC human liver single exon encoded peptides of the invention.

CC Note: The sequence information for this patent does not appear in the

CC printed specification but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 190 AA;

Query Match 20.1%; Score 1005; DB 22; Length 190;

Best Local Similarity 97.4%; Pred. No. 2e-63;

Matches 185; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 753 KRPHCCGCGCKAFKFKHHLIHSRLHSGEKPYQCDKCGKRFSGSHSGYOHMHRYSYCKR 812

Db 1 KRPHCCGCGCKAFKFKHHLIHSRLHSGEKPYQCDKCGKRFSGSHSGYOHMHRYSYCKR 60

Qy 813 EAEREAERAREKXHLGPTLIMNRAVLOSITPGYSDSEERSEMPRDGSEKHEKE 872

Db 61 EAEREAERAREKXHLGPTLIMNRAVLOSITPGYSDSEERSEMPRDGSEKHEKE 120

Qy 873 GEQYGKLRDRDGEDEESESSEKNSMDTPTIRDEETGHSMDSSGDKMETKS 932

Db 121 GEDYGKLRDRDGEDEESESSEKNSMDTPTIRDEETGHSMDSSGDKMETKS 180

Qy 933 DHEEDNMEDG 942

Db 181 DHEEDNMEDG 190

RESULT 9

ABG51171

ID ABG51171 standard; Peptide; 190 AA.

XX

AC ABG51171;

XX

DT 25-FEB-2003 (first entry)

XX

DE Human liver peptide, SEQ ID No 29819.

XX

KM Human, liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;

KM hypercholesterolaemia; coronary heart disease.

XX

OS Homo sapiens.

XX

PN WO200157273-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00664.

XX

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488898/53.

XX

PT Human genome-derived single exon nucleic acid probes useful for

XX analysing gene expression in human adult liver -

PS Claim 27; SEQ ID No 29819; 658bp; English.

XX

CC The invention relates to a single exon nucleic acid probe (SENP) (1) for

CC measuring human gene expression in a sample derived from human adult

CC liver, comprising one of 13109 defined nucleotide sequences given in the

CC specification (or complements/ fragments). The probe hybridises at high

CC stringency to a nucleic acid molecule expressed in the human adult

CC liver. (1) may be used for predicting, measuring and displaying gene

CC expression in samples derived from human adult liver. The genes

CC identified may be involved in genetic liver diseases such as cirrhosis,

CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which

CC is associated with coronary heart disease. ABG47348-ABG5930 represent

CC human liver single exon encoded peptides of the invention.

CC Note: The sequence information for this patent does not appear in the

CC printed specification but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 190 AA;

Query Match 20.1%; Score 1005; DB 22; Length 190;

Best Local Similarity 97.4%; Pred. No. 2e-63;

Matches 185; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY	753	KRHQCOICKKAFKXKHGHLIERSRLHSGKPKQCCCKGRFSHSGSYGQHNHNYCYCKR	813
Db	1	KRHQCOICKKAFKXKHGHLIERSRLHSGKPKQCCCKGRFSHSGSYGQHNHNYCYCKR	60
QY	813	EAEREAAERAREAREKCHLGPTELLMNRAYLGSIIPQGYSDSERBSMPRODESEKEHE	872
Db	61	EAEREAAERAREAREKCHLGPTELLMNRAYLGSIIPQGYSDSERBSMPRODESEKEHE	120
QY	873	GEGYCYKTLARRDDDEEEEEESENKSMDDPTETIRPEETGCHSMDSDSDCKMTKS	932
Db	121	GEGYCYKTLARRDDDEEEEEESENKSMDDPTETIRPEETGCHSMDSDSDCKMTKS	180
QY	933	DHEEDNMEDG 942	
Db	181	DHEEDNMEDG 190	

RESULT 10
 ABB28955
 ID ABB28955 standard; Peptide; 190 AA.
 AC ABB28955;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Peptide #1606 encoded by breast cell single exon nucleic acid probe.
 XX
 KM Human; microarray; single exon probe; gene expression; breast;
 KM disease; cancer.
 XX
 OS Homo sapiens.
 PN W0200157271-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001MO-US00662.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-496933/54.
 XX
 PT New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -
 XX
 PS Claim 27; SEQ ID NO 11923; 327bp + sequence listing; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting
 CC the probes with a collection of detectably labeled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a peptide encoded by a single exon nucleic acid

Query Match	20.1%	Score 1005;	DB 22;	Length 190;
Best Local Similarity	97.4%	Pred. No. 2e-63;		
Matches 185;	Conservative 2;	Mismatches 3;	Indels 0;	Gaps 0
QY	753	KRPHQOICCKAFYGHKHLIEHSRLHSGEKPYQCDKCGKRFSGSGYSQHNNHRYSYCKR	812	
Db	1	KRPHQOICCKAFYGHKHLIEHSRLHSGEKPYQCDKCGKRFSGSGYSQHNNHRYSYCKR	60	
QY	813	EAEREPAERERARKGHGLPTLELMNRAVYLSITPPQGSDEERESMPRDGSEKEHEKE	872	
Db	61	EAEREPAERERARKGHGLPTLELMNRAVYLSITPPQGSDEERESMPRDGSEKEHEKE	120	
QY	873	GEGYGYGLRRRDGDEEEEEESEENSMQDPTPIREESTGDSMDSSSDGDKMETKS	932	
Db	121	GEGYGYGLRRRDGDEEEEEESEENSMQDPTPIREESTGDSMDSSSDGDKMETKS	180	
QY	933	DHEEDNMEDG 942		
Db	181	DHEEDNMEDG 190		
RESULT 11				
ABB31134				
ID	ABB31134	standard; peptide: 190 AA.		
AC	ABB31134;			
DT	01-FEB-2002	(first entry)		
XX	Peptide #3785	encoded by breast cell single exon nucleic acid probe.		
DE	Human; microarray; single exon probe; gene expression; breast; disease; cancer.			
KW	Homo sapiens.			
XX	WO200157271-A2.			
PN	09-AUG-2001.			
XX	30-JAN-2001; 2001WO-US00662.			
PF	04-FEB-2000; 2000US-0180312.			
XX	26-MAY-2000; 2000US-0207456.			
PR	30-JUN-2000; 2000US-0608408.			
XX	03-AUG-2000; 2000US-0632366.			
PR	21-SEP-2000; 2000US-0234687.			
XX	27-SEP-2000; 2000US-0236359.			
PR	04-OCT-2000; 2000GB-0024263.			
XX	{MOLE-} MOLECULAR DYNAMICS INC.			
XX	Penn SG, Hanzel DK, Chen W, Rank DR;			
XX	WPI; 2001-496933/54.			
DR	New spatially-addressable set of single exon nucleic acid probes,			
PT	useful for measuring gene expression in sample derived from human breast, comprises number of single exon nucleic acid probes			
PT	Claim 27; SEQ ID NO 14102; 327bp + sequence listing; English.			
PS	The invention relates to a spatially-addressable set of single exon			
CC	nucleic acid probes for measuring gene expression in a sample derived			
CC	from human breast and BT 474 cells. The method involves contacting			
CC	the probes with a collection of detectably labeled nucleic acids			

derived from mRNA of human breast, and then measuring the label bound to each probe of the microarray. The probes are useful for verifying the expression of regions of genomic DNA predicted to encode proteins. They are useful for gene discovery, and for determining predilection and/or prognosing breast disease. Gene expression analysis is useful for assessing the toxicity of chemical agents on cells. The microarray of this invention presents a far greater diversity of probes for measuring gene expression, with far less bias than expressed sequence tag microarrays. The method is suitable for rapid production of functional information from genomic sequence. The present sequence is a peptide encoded by a single exon nucleic acid probe of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from www.ncbi.nlm.nih.gov/pubmed/125000000.

SQ **Sequence** **190 AA;**

Query Match	20.1%	Score 1005	DB 22	Length 190
Best Local Similarity	97.4%	Pred. No. 2e-63		
Matches 185	Conservative	2	Mismatches 3	Indels 0
				Gaps 0

Qy	753	KRPHOCIOCKKA	PKKHGHIL	IEHSRLHSGEK	PCQCDCKGR	SRSHSGSY	QJHNNHRY	YSYCKR	812
Db	1	KRPHOCIOCKKA	PKKHGHIL	IEHSRLHSGEK	PCQCDCKGR	SRSHSGSY	QJHNNHRY	YSYCKR	60
Qy	813	EAEREAEAREARE	KHGLGPTBL	LNMRAYLOSIT	POGYSDESE	RESMPRODESE	KEKEKE		872
Db	61	EAEREAEAREARE	KHGLGPTBL	LNMRAYLOSIT	POGYSDESE	RESMPRODESE	KEKEKE		120
Qy	873	GEEGYCKLRRD	DEEEEEEEN	KKMDTPETIR	DEEETGHS	MDSSGSDG	KMETKS		932
Db	121	GEEGYCKLRRD	DEEEEEEEN	KKMDTPETIR	DEEETGHS	MDSSGSDG	KMETKS		180
Qy	933	DHEEDNMDG	942						
Db	181	DHEEDNMDG	190						

RESULT 12

ABB34123
ID ABB34123 standard; Peptide; 190 AA.

AC ABB34123;

DT 04-FEB-2002 (first entry)

DE Peptide #1629 encoded by human foetal liver single exon probe.

KW Human; foetal liver; gene expression; single exon nucleic acid probe.

OS Homo sapiens.

PN W0200157277-A2.

PD 09-AUG-2001

PF 30-JAN-2001; 2001WO-US00669.

PR	04-FEB-2000	2000US-0180312
PR	26-MAY-2000	2000US-0207456
PR	30-JUN-2000	2000US-0508048
PR	03-AUG-2000	2000US-0632366
PR	21-SEP-2000	2000US-0234687
PR	27-SEP-2000	2000US-0236359
PR	04-OCT-2000	2000GE-0024263

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483447/52.

XX

PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 27: SEQ ID NO 26758: 639pp + sequence listing; English.
PS

PS Claim 27; SEQ ID NO 26758; 639pp + sequence listing; English

CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at fp.wipo.int/pub/published_pct_sequences.

SQ Sequence 190 AA;

Query Match	20.1%	Score 1005	DB 22	Length 190
Best Local Similarity	97.4%	Pred. No. 2e-63		
Matches 185; Conservative	2	Mismatches	3	Indels 0; Gaps 0

QY	75.5	KKPHOOICCKKAFKKHKLIEHSRLHSEKPYOCDCKGKPSHGSGYSQHMNHRYSYCKR	81.2
Db	1	KKPHOOICCKKAFKKHKLIEHSRLHSEKPYOCDCKGKPSHGSGYSQHMNHRYSYCKR	60
QY	81.3	EAEEREAERAEARKGHGPTLLMNRVYLOSITPQGYSDSEERESNPRDGESEKHEKE	87.2
Db	61	EAEEREAERAEARKGHGPTLLMNRVYLOSITPQGYSDSEERESNPRDGESEKHEKE	120
QY	87.3	GEBGKGLRRDDGDEEBEESBESNKSMDTPETTRDEEETGHSMDSSSEDKMETKS	93.2
Db	121	GEBGKGLRRDDGDEEBEESBESNKSMDTPETTRDEEETGHSMDSSSEDKMETKS	180
QY	93.3	DHEEDNMDG 942	
Db	181	DHEEDNMDG 190	

RESULT 13

ABB36333
ID ABB36333 standard; Peptide; 190 AA.

AC ABB36333;

DT 04-FEB-2002 (first entry)

Peptide #3839 encoded by human foetal liver single exon probe.

KW Human; foetal liver; gene expression; single exon nucleic acid probe.

OS Homo sapiens.

PN WO200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US006669.

PR	04-FEB-2000	2000US-0180312
PR	26-MAY-2000	2000US-0207456
PR	30-JUN-2000	2000US-0608408
PR	03-AUG-2000	2000US-0632366
PR	21-SEP-2000	2000US-0234687
PR	27-SEP-2000	2000US-0236359
PR	04-OCT-2000	2000GB-0024263

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483447/52.

XX

PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human fetal liver -
 XX
 PS Claim 27; SEQ ID NO 28968; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 190 AA;
 XX
 Query Match 20.1%; Score 1005; DB 22; Length 190;
 Best Local Similarity 97.4%; Pred. No. 2e-63;
 Matches 185; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 753 KRPHQCQICKKAFKXKHLIEHSRLHSGEKPYQCDKCGKRFSGSYSGHMHRYSYCKR 812
 Db 1 KRPHQCQICKKAFKXKHLIEHSRLHSGEKPYQCDKCGKRFSGSYSGHMHRYSYCKR 60
 QY 813 EAEREERAEAREKRGHLEPTLLMNRAYLQSTTPQGYSDSERESMPRDGSEKEHEKE 872
 Db 61 EAEREERAEAREKRGHLEPTLLMNRAYLQSTTPQGYSDSERESMPRDGSEKEHEKE 120.
 QY 873 GEEGYGKLRRDGDEEEEEESENKSMQDTPETIRDEETGDSMDSDSGKMETKS 932
 Db 121 GEEGYGKLRQDDEFEFEFESENKSMQDTPETIRDEETGDSMDSDSGKMETKS 180
 QY 933 DHEEDNMEDG 942
 Db 181 DHEEDNMEDG 190
 RESULT 14
 ABB19564
 ID ABB19564 standard; Protein; 190 AA.
 XX
 AC ABB19564;
 DT 23-JAN-2002 (first entry)
 XX
 DE Protein #1563 encoded by probe for measuring heart cell gene expression.
 XX
 KM Human; gene expression; heart; microarray; vascular system;
 KM cardiovascular disease; hypertension; cardiac arrhythmia;
 KM congenital heart disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200157274-A2.
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00666.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488899/53.
 XX

PT Single exon nucleic acid probes for analyzing gene expression in human
 XX hearts -
 PS Claim 15; SEQ ID NO 21334; 530pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart (see
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
 CC probe. The probes may be used for predicting, measuring and displaying
 CC gene expression in samples derived from the human heart via microarrays.
 CC By measuring gene expression, the probes are useful for predicting,
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the
 CC human heart and vascular system e.g. cardiovascular disease,
 CC hypertension, cardiac arrhythmias and congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 190 AA;
 XX
 Query Match 20.1%; Score 1005; DB 22; Length 190;
 Best Local Similarity 97.4%; Pred. No. 2e-63;
 Matches 185; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 753 KRPHQCQICKKAFKXKHLIEHSRLHSGEKPYQCDKCGKRFSGSYSGHMHRYSYCKR 812
 Db 1 KRPHQCQICKKAFKXKHLIEHSRLHSGEKPYQCDKCGKRFSGSYSGHMHRYSYCKR 60
 QY 813 EAEREERAEAREKRGHLEPTLLMNRAYLQSTTPQGYSDSERESMPRDGSEKEHEKE 872
 Db 61 EAEREERAEAREKRGHLEPTLLMNRAYLQSTTPQGYSDSERESMPRDGSEKEHEKE 120
 QY 873 GEEGYGKLRRDGDEEEEEESENKSMQDTPETIRDEETGDSMDSDSGKMETKS 932
 Db 121 GEEGYGKLRQDDEFEFEFESENKSMQDTPETIRDEETGDSMDSDSGKMETKS 180
 QY 933 DHEEDNMEDG 942
 Db 181 DHEEDNMEDG 190
 RESULT 15
 ABB21690
 ID ABB21690 standard; Protein; 190 AA.
 XX
 AC ABB21690;
 DT 23-JAN-2002 (first entry)
 XX
 DE Protein #3689 encoded by probe for measuring heart cell gene expression.
 XX
 KM Human; gene expression; heart; microarray; vascular system;
 KM cardiovascular disease; hypertension; cardiac arrhythmia;
 KM congenital heart disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200157274-A2.
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00666.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488899/53.
 XX

THIS PAGE BLANK (USP 10)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comphen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2004, 16:27:45 ; Search time 28 Seconds
(without alignments)
3242.258 Million cell updates/sec

Title: US-09-964-238-2

Perfect score: 4991
Sequence: 1 MLTQAGNRKPKCTECGKAF.....DGKMETKSDHEEDMEDMG 944

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	4273	85.6	1214	2	JC7259 Smad interacting p
2	1665	33.4	1114	2	deltae1 - chicken
3	1657.5	33.2	1124	2	zinc finger protei
4	1605.5	32.2	1154	2	E-box-binding repr
5	1599	32.0	1043	2	DNA-binding protei
6	1587.5	31.8	1117	2	delta-crystallin/E
7	819.5	16.4	1060	2	homeotic protein z
8	741	14.8	880	2	hypothetical prote
9	379	7.6	1191	2	zinc finger protei
10	358	7.2	1350	2	finger protein - A
11	356.5	7.1	1042	2	EVII protein - hum
12	355.5	7.1	1051	2	oncogene Evi-1 - h
13	345	6.9	1042	2	transcription regu
14	328.5	6.6	701	2	hypothetical prote
15	315.5	6.3	803	2	zinc finger protei
16	315.5	6.1	469	2	zinc finger protei
17	302.5	6.1	1173	2	proctk2 - chicken
18	299.5	6.0	1615	2	ras-responsive ele
19	298.5	6.0	1186	2	O/E-1-associated z
20	296	5.9	580	2	specmatogenesis pr
21	286	5.9	675	2	zinc-finger protei
22	292	5.9	651	2	finger protein 2,
23	290.5	5.8	1350	2	finger finger protei
24	289.5	5.8	615	2	finger protein (cl
25	289.5	5.8	1323	2	spalt protein - mo
26	285.5	5.7	636	2	gene NK10 protein
27	285	5.7	460	2	hypothetical prote
28	285	5.7	856	2	B-lymphocyte-induc
29	284	5.7	710	2	zinc finger protei

30	280.5	5.6	511	2	IS0114	early growth respo
31	279	5.6	644	2	T46277	hypothetical prote
32	278.5	5.6	693	2	I37570	zinc finger protei
33	277.5	5.6	536	2	S06548	finger protein (cl
34	276	5.5	553	2	S22954	finger protein zfp
35	275.5	5.5	706	2	A48752	B-cell CLL/Lymphom
36	274.5	5.5	508	2	A32225	nerve growth facto
37	274.5	5.5	789	2	A39564	transcription repr
38	273.5	5.5	727	2	B60191	transcription regu
39	272.5	5.5	543	2	A41211	early growth respo
40	271.5	5.4	706	2	I52586	B-cell CLL/Lymphom
41	270	5.4	533	2	U50304	developmental cont
42	266	5.3	668	2	T08725	probable finger pr
43	265	5.3	733	2	JC7679	dendritic cell-der
44	264.5	5.3	678	2	S62939	hypothetical prote
45	264	5.3	707	2	S68858	finger protein - m

ALIGNMENTS

RESULT 1

Smad interacting protein 1 - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C:Accession: JC7259
R:Eisaki, A.; Kuroda, H.; Fukui, A.; Aashima, M.
Biochem. Biophys. Res. Commun. 271, 151-157, 2000
A:Title: XSLP1, a member of two-handed zinc finger proteins, induced anterior neural mar
A:Reference number: JC7259
A:Accession: JC7259
A:Molecule type: mRNA
A:Residues: 1-1214 <EIS>
A:Experimental source: egg
C:Comment: This protein, a two-handed zinc finger and a homeodomain protein, is a transci
signal pathway. It plays the roles in early neural development.
C:Genetics:
A:Gene: xslp1
C:Keywords: egg; zinc finger

Query Match	85.6%	Score 4273	DB 2	Length 1214	Best Local Similarity 86.1%	Pred. No. 1.4e-215	Matches 815	Conservative 56	Mismatches 68	Indels 8	Gaps 5
QY	1	MLTQAGNRKPKCTECGKAFKXHLKEHLRHSGEKYECPNCKKRSFGSYSHISS	60								
DB	272	MLTQAGNRKPKCTECGKAFKXHLKEHLRHSGEKYECPNCKKRSFGSYSHISS	331								
QY	61	KKICGLISVNGRMNNITGSSPNSVSSPTNSAITQLRNKLKLGKPLMSSEOTGLKIK	120								
DB	332	KKICGLISVNGRMNNITGSSPNSVSSPTNSAITQLRNKLKLGKPLMSSEOTGLKIK	391								
QY	121	TEPLDFNDYKYLW-ATHGPGSSGSPMNGLGATSPBVGHPAQSQPMQHLGYMEAPLIGF	179								
DB	392	TESLDYNDYKYLW-ATHGPGSSGSPMNGLGATSPBVGHPAQSQPMQHLGYMEAPLIGF	451								
QY	180	PTMNSNLSEVQKVLQYVNTVYSRQMDCKTEDISLKKYHMKDPDSQDEBQVTSFNP	239								
DB	452	PTMNSNLSEVQKVLQYVNTVYSRQMDCKTEDISLKKYHMKDPDSQDEBQVTSFNP	511								
QY	240	VGLPVSHNGATKSIIDTLEKVNFAKACLOSLFTDSRQISNKKKGLRTLIDVTDK	299								
DB	512	VGLPVSHNGATKSIIDTLEKVNFAKACLOSLFTDSRQISNKKKGLRTLIDVTDK	571								
QY	300	MIENHSISTPSCQCKSPFGPIPLQHERRYLCMNEIRKAVIQPHENIVPNKAGVFD	359								
DB	572	MIENHSISTPSCQCKSPFGPIPLQHERRYLCMNEIRKAVIQPHENIVPNKAGVFD	631								
QY	360	NKALLISVLSKGLTSPINPYKDMSTLKAYVAMNMPNSDELKLSIAVGLQDFEYKE	419								
DB	632	KOALLISVLSKGLTSPINPYKDMSTLKAYVAMNMPNSDELKLSIAVGLQDFEYKE	691								

Qy	420	WFEORXVYQVNSNSPSLERTS--KPLAPNSNPPTXOSLIPRSVVKMDSTTSSIMEH	477
Db	692	WFEORXVYQVNSNSPSLERTSMAELATILANTPTKOS--ARSPKXVDPTTSSIMEH	749
Qy	478	NSVTSQDBPLRLTYKSHFTNIKAV-DKLDHSRSTNPSPNLNSTSSKXSHSSSYTPNSFS	536
Db	750	NRVNCNDPPLRLTXSNHPSAKMPPVLDKLDHSRSTNPSPNLNSTSSKXSHSSSYTPNSFS	809
Qy	537	SEEOAEPBLDLSLKOMEPEKGIATKUKTATYSINLDHNSVSSSENSDEPPLRLTYIKK	596
Db	810	SEEOAEPBLDLSVFKLANESTKIITATKXKSAPNNIYDHSVSLSSSEVDEPPLRLTYIKK	869
Qy	597	EFSNSNNLDKSNPNVFCGNPFSAPKPLTYPLPPOSAPPATMPDPVOTSIPGLRYPGLD	656
Db	870	EFCNANM--DKSTSPPLFGILNPFSGKPLYSALRPOSAPPATMPDPVOTGIPGLRSYGLD	927
Qy	657	QMSFLPHAYIYPPGAATFADMOORXKQORQOGQGLDLDGADYMGSLDMTDSCL	716
Db	928	QMSFLPHAYIYPNGAATFADMOORRKYORKQOGQGLDLDGTQYMGSLDMTDSCLS	987
Qy	717	RKKIKTESGMYACDLCDKTFOKSSSLLRHKYEHNGKAPHOQOJCKKAFKIKHLLIHSR	776
Db	988	RKKIKTESGMYACDLCDKTFOKSSSLLRHKYEHNGKAPHOQOJCKKAFKIKHLLIHSR	1047
Qy	777	LHSGEKPQOCCKGGRFSGSSYQOMHNRKSYCKREAEEREAERERAREKNGHLPBELL	836
Db	1048	LHSGEKPQOCCKGGRFSGSSYQOMHNRKSYCKREAEEREAERERAREKNGHLPBELL	1107
Qy	837	MNRAYLOSITPQGSYDSSERESMPDGESEKHEHEKGEYKGLRRRDGDEEEEBEES	896
Db	1108	MNRAYLOSITPQGSYDSSERESMPDRGRLEHEHEGDDYVDKLRQYGDDEEFEEBEES	1167
Qy	897	ENKSMDDTPETIIRDEEETGDSHMDSSSDGCKMETKSDHEEONMEDGM	943
Db	1168	ENKSMDDTPETIIRDEEENGDSHMDSSSDGCKMEAKSDHEEIEIMDGM	1214

```

RESULT 2
150222
deltaEF1 - chicken
C:Species: Gallus gallus (chicken)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: 150222
R:Funahashi, J.; Sekido, R.; Murai, K.; Kamachi, Y.; Kondoh, H.
Development_119, 433-446, 1993
A:Title: Delta-crystallin enhancer binding protein delta EF1 is a zinc finger-homeodomain
A:Reference number: 150222; PMID:94116444; PMID:7904558
A:Accession: 150222
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1114 <FON>
A:Cross-references: GB:D14313; NID:G391631; PIDN:BA003259.1; PID:G391632

Query Match      33.4% Score 1665; DB 2; Length 1114;
Best Local Similarity 41.3% Pred. No.2.7e-79;
Matches 394; Conservative 151; Mismatches 296; Indels 112; Gaps 26;

Oy      2 LTQAGNKKFKCTECGKAFKYKHLKEHLRIHSGEPCECNCKKRFSSHSYSYSHISK 61
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      231 VTQSSGNKKFKCTECGKAFKYKHLKEHLRIHSGKPEECNCKKRFSSHSYSYSHISK 290

Oy      62 KCIGLISVNGRMNNIKTG--SSPNSVSSSPTNSAITGLRNKLENGKPLSMSEQTGLIKI 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      291 KCIGLMPVKGRARSGLKTSQCSSP-SLSASPGSPARPQIRQIKEN-KPL--QEDLPVNVQI 346

Oy      120 KTEPLDFNDYKVLMTHTFGSSGSSPFMNGGLGATSPLVGHPSAQSPMQL---GVGMEAPL 176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      347 KTEPVDY-EFKFIVVASGINCSTPLQNGVFSGSGSPLDQTSSPQGVAVVLPVPTGLVSP 405

Oy      177 LGFPMNNSISVQVVLQI-VNNTYSRQMKDCKTDISKLGYHMKDKCSQPEEQGYTSP 235
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      406 -----SINLSIQVLAVDGAVIRQVLENNHNLASKS---QETISNASIQDAGHS 455

```

[illegible]

```

RESULT 3
JX0293
zinc finger protein AREB6 - human
C.Species: Homo sapiens (man)
C.Date: 03-Feb-1994 #sequence_revision 20-Aug-1994 #ext_change 28-May-1999
C.Accession: JX0293
R.Watanabe, Y.; Kawakami, K.; Hirayama, Y.; Nagano, K.
J. Biochem. 114, 849-855, 1993
J.Title: Transcription factors positively and negatively regulating the Na,K-ATPase alpha
A.Reference number: JX0293; MUID:94186507; PMID:8138542
A.Accession: JX0293
A.Molecule type: mRNA
A.Residues: 1-1124 <MAT>
A.Cross-references: GB:DI5050; NID:9457560; PIDD:BA03646.1; PID:d1004160; PID:9558522
C.Keywords: zinc finger
F.117-193/Region: zinc finger
F.202-222/Region: zinc finger
F.242-262/Region: zinc finger
F.270-292/Region: zinc finger
F.306-326/Region: zinc finger
F.334-354/Region: zinc finger
F.362-381/Region: zinc finger
F.389-412/Region: glutamic acid-rich

```


Query Match 33.2%; Score 1657.5; DB 2; Length 1124;
 Best Local Similarity 41.9%; Pred. No. 6,8e-79;
 Matches 406; Conservative 147; Mismatches 308; Indels 107; Gaps 32;

```

2 LTQGNKRFKTECGKAFKTKHLEHRIHSGEKPCEPKRFRSHSGSYSHISSK 61
21 VTQSCNKKFKTECGKAFKTKHLEHRIHSGEKPCEPKRFRSHSGSYSHISSK 290
62 KICGLISVNGRMNNIKTG--SSPNSVSSPTNSAITQLRNKLNGKPLMSBQTLKI 119
291 KCISLIIPVNGRPRIGTKITSQCSSP-SLSASPGSPTRPQRIKTIEN-KPL--QEOLSVNOI 346
120 KTEPLDNDYKVLMAATHFGSSGSPFMNGGLGATSPGVHPSAQSPMOHL--GVGMEAPL 176
347 KTEPDYD--EFKPIVVASGINCSTPLONGVFTGGGPLQATSSPQGVAVVLPVGLVSPDI 405
177 LGFPMNSNLSEVQVLOI-VDNTYSRQKMDCKTEBISLKGYNHMKDPCSOBEGCVTSP 235
406 -----SINLSDIQVNLKAVDGNVIRQYLE---NNQANLASKEOETINASPIQGGSHSV 456
236 NIPVGLPVVSHNGATKSIIDYTLKEVNEAKACLOSLTDSRRQISNIKEKRLTLDIV 295
457 -ISALSPLVODDGTTKIITINSLEQPSQLOVPQNLKKNPVATNSCKSEKLPEDLTVK 515
236 TD-DKMEINH-SISTPFSQCFKESFPGP-PLQHHERYLCRMNEIKAVILOPHENIVPN 352
516 SEKDSFEGGVNDSTCLCDDC---PGDINALPELKHV-----DLKQTPQ-----PP 560
353 KAGVVDNKKALLSSVLSSEKGLTSPINPYKDHMSVTKAYYAMNPNDELKISIANVL 412
561 LPAAEAKEPSSVSATGDNLSPPSQPLKNLISLTKAYVALNAPSAEELSKIDSVNL 620
413 POEFVKEPFEORKVYQVNSRSPSLERTSKPLAPNSNPTTKOSLLPRSPVKPMDSITSPS 472
621 PLDVKKAFKFKQAGQIS-----VQSSBSPSPGKV-----NIPAKNNDQPSGAN 666
473 IAEIHNVSVCDDPRLRLTKSSHFTNIKAVDKLHRSNTPSPPLNLSSTSSKNSHSSSYTP 532
667 ANEPQDSTVNLQSPKMTNS---PVLPGVSTTNGSSSTPSPPLNLSSTSSNTQGYLYTA 723
533 NSFSEBELOAEPLDLSLPKQKREPKCIATKNTKATKATSLNDHNSVSSSENS-----DE 587
724 EG-AQEEBPQVEPLDLSLPKQGE-----LERSSTITVYQNSVSVQBE 766
588 PLNLPIKKEPFSNNLNDKSNPNVGMNPFSAKPLTYPLPQSAFPATFMPVQTSIP 647
767 PLNLSCAKKEPQKDCVTD--SEPVVNVLPSPANPINIAIPVTYQLPITVAIADQNSVP 824
648 GLRPPYGLDQMSFLPHMAATYPTGATFADMOQRKYQKQGFQGLDGAODYMSGLDD 707
825 CLRALAANKQITLIPQVATYTTTSP-AYQEPPLKVIQPNQNOBERODTSSBGSVNVED 883
708 MTDSDSCLSRKIKITKTESGMVACDLCDKTFQKSSSLRHKYHTGKRPQCOICCKAFK 767
884 QNDSDSTPCKKRRKTKENGMVACDLCDKIFQKSSSLRHKYHTGKRPHECGICCKAFK 943
768 KHLLEHSHLSGKRYQCDKCGKRFSSHSGSYSOHNNHYSYCKRAEEREAERARX 827
944 KHLLEHSHLSGKRYQCDKCGKRFSSHSGSYSOHNNHYSYCKRAEEREAERARX 999
828 GHIGPTELMNRAVLOSITP-QYSDSEERESMPR--DGESEKHEKEE-----GYK 879
1000 -BAGP-EILSNHVGARASPSQG--DSDERESLTREDEDSKEKEBEKEKEMBEIQEKE 1055
880 LRRDGDDEEBEESSEKSMQDTPETLRDEBETGDHMSDSSD-----GKMET 930
1056 CKRPQDDEEBEESSEKSMQDTPETLRDEBETGDHMSDSSD-----GKMET 930
931 KSDHEEDN 938
1115 QVSEKTN 1122

```

RESULT 4

A56242

E-box-binding repressor ZEB - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 07-May-1999

C/Accession: A56242

R/Gene: 14, 6153-6163, 1994

Mol. Cell. Biol. 14, 6153-6163, 1994

A>Title: Displacement of an E-box-binding repressor by basic helix-loop-helix proteins;

A/Reference number: A56242; MUID:94344126; PMID:8065348

A/Accession: A56242

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-1154 <GEN>

A/Cross-references: GB:U19669

C/Keywords: DNA binding; zinc finger

Query Match 33.2%; Score 1605.5; DB 2; Length 1154;
 Best Local Similarity 41.1%; Pred. No. 3.7e-76;
 Matches 399; Conservative 142; Mismatches 314; Indels 115; Gaps 32;

```

2 LTQGNKRFKTECGKAFKTKHLEHRIHSGEKPCEPKRFRSHSGSYSHISSK 61
25 VTQSCNKKFKTECGKAFKTKHLEHRIHSGEKPCEPKRFRSHSGSYSHISSK 324
62 KICGLISVNGRMNNIKTG--SSPNSVSSPTNSAITQLRNKLNGKPLMSBQTLKI 119
325 KCISLIIPVNGRPRIGTKITSQCSSP-SLSASPGSPTRPQRIKTIEN-KPL--QEOLSVNOI 380
120 KTEPLDNDYKVLMAATHFGSSGSPFMNGGLGATSPGVHPSAQSPMOHL--GVGMEAPL 176
381 KTEPDYD--EFKPIVVASGINCSTPLONGVFTGGGPLQATSSPQGVAVVLPVGLVSPDI 439
177 LGFPMNSNLSEVQVLOI-VDNTYSRQKMDCKTEBISLKGYNHMKDPCSOBEGCVTSP 235
440 -----SINLSDIQVNLKAVDGNVIRQYLE---NNQANLASKEOETINASPIQGGSHSV 490
236 NIPVGLPVVSHNGATKSIIDYTLKEVNEAKACLOSLTDSRRQISNIKEKRLTLDIV 295
441 -ISALSPLVODDGTTKIITINSLEQPSQLOVPQNLKKNPVATNSCKSEKLPEDLTVK 549
491 -ISALSPLVODDGTTKIITINSLEQPSQLOVPQNLKKNPVATNSCKSEKLPEDLTVK 549
236 TD-DKMEINH-SISTPFSQCFKESFPGP-PLQHHERYLCRMNEIKAVILOPHENIVPN 352
550 SEKDSFEGGVNDSTCLCDDC---PGDINALPELKHV-----DLKQTPQ-----PP 594
353 KAGVVDNKKALLSSVLSSEKGLTSPINPYKDHMSVTKAYYAMNPNDELKISIANVL 412
595 LPAAEAKEPSSVSATGDNLSPPSQPLKNLISLTKAYVALNAPSAEELSKIDSVNL 654
413 POEFVKEPFEORKVYQVNSRSPSLERTSKPLAPNSNPTTKOSLLPRSPVKPMDSITSPS 472
655 PLDVKKAFKFKQAGQIS-----VQSSBSPSPGKV-----NIPAKNNDQPSGAN 700
473 IAEIHNVSVCDDPRLRLTKSSHFTNIKAVDKLHRSNTPSPPLNLSSTSSKNSHSSSYTP 532
701 ANEPQDSTVNLQSPKMTNS---PVLPGVSTTNGSSSTPSPPLNLSSTSSNTQGYLYTA 757
533 NSFSEBELOAEPLDLSLPKQKRE-----PKGIATKNTKATKATSLNDHNSVSSSENS 585
758 EG-AQEEBPQVEPLDLSLPKQGEELERILPPEQCLFCBGR----- 797
586 DEPLNLTKKEPFSNNLNDKSNPNVGMNPFSAKPLTYPLPQSAFPATFMPVQTS 645
798 ---LEVLAKKEPQKDCVTD--SEPVVNVLPSPANPINIAIPVTYQLPITVAIADQNS 852
646 IPGLRPPYGLDQMSFLPHMAATYPTGATFADMOQRKYQKQGFQGLDGAODYMSGL 705
853 VPCRALAANKQITLIPQVATYTTTSP-AYQEPPLKVIQPNQNOBERODTSSBGSVNV 911
706 DMTDSDSCLSRKIKITKTESGMVACDLCDKTFQKSSSLRHKYHTGKRPQCOICCKAF 765
912 BDQNDSDSTPCKKRRKTKENGMVACDLCDKIFQKSSSLRHKYHTGKRPHECGICCKAF 971

```

```

QY 766 KKHKHLIHSRLHSGEKPYQCDCKGRFSGSHSGSYSGHMHRYSYCKREAEREAREAR 825
D 972 KHKHHLIHSRLHSGEKPYQCDCKGRFSGSHSGSYSGHMHRYSYCKREAEREAREAR 1029
QY 826 EKHLGTELLMNRAYIQSTP-QGYSDSBERESMPR--DGESEKHEKEGEE-----GY 877
D 1030 --EAGP-ELTSMNHVGARASPSQG--DSDERESLTREDEDSKKEEEDKEMELQEE 1083
QY 878 GKLRRRDDEEESESESESESESESESESESESESESESESESESESESESESESESE 928
D 1084 KECEKPGQDEEESESESESESESESESESESESESESESESESESESESESESESE 1142
QY 929 ETKSDHEEDN 938
D 1143 SEQVSEKTN 1152

```

RESULT 5

A56037
DNA-binding protein B2P - golden hamster
C/Species: Mesocricetus auratus (golden hamster)
C/Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 05-Nov-1999
C/Accession: A56037
R/Franklin, A.J.; Jecton, T.L.; Shelton, K.D.; Magnuson, M.A.
Mol. Cell. Biol. 14, 6773-6788, 1994
A/Title: B2P, a novel serum-responsible zinc finger protein that inhibits gene transcription
A/Reference number: A56037; MUID:95021206; PMID:7935395
A/Accession: A56037
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1043 <PRA>
A/Cross-references: GB:L13856; NID:g437340; PIDN:AAC37667.1; PID:g437341

Query Match 32.0%; Score 1599; DB 2; Length 1043;
Best Local Similarity 40.2%; Pred. No.7e-76;
Matches 390; Conservative 142; Mismatches 308; Indels 130; Gaps 30;

```

QY 2 LTQAGNRKFKCTECGAKFYKHLKHLRIHSGEKPYECPCNCKRFSHSGSYSHISK 61
D 155 VTQSGNRKFKCTECGAKFYKHLKHLRIHSGEKPYECPCNCKRFSHSGSYSHISK 214
QY 62 KCTGLISVNGRNRNNTTGGSPN-SVSSSPNNSATITOLRNKLENGKRLSSEOTGLIKT 120
D 215 KCTSLMPVNGRPRSGLTQSCPSPSLSASPSPTRPQIRQKIEN-KPL--QEPUSVNOIK 271
QY 121 TEPLDFNDYKVLMTATHGFSGSSPFMNG---GLG---ATSPGVHPSAGSPMCHLGVME 173
D 272 TEPLDFY-EPKPIVVASGINCSTPLQNGVFGSGGLQATSSFGGVVAVLTPT---VGLV 326
QY 174 APPLGPTMNSNLSEVQKVLQI-VDNTVSRQMDCKTEDISKLGKYMKPCSQPEQGV 232
D 327 SPI-----SINSDIQNLKVALDGNVIRQVLENNQASL-----ASKGEAV 368
QY 233 TSPN-----TPPVGLPVVSHNGATKSIIDYLEVNEAKACQLGTLTDSRRQSNIK 284
D 369 NASIIQCGSHVSISALPLVDQGTGKIINYSLEQPSQLQVVPQWLKTEHPAPTSCK 428
QY 285 KEKRLTILIDVTDKMLNHSISIRPSCQCFKESFPPIPLHQHERLTCMMNERIKAVLQ 344
D 429 SEKLPE--DLTVSEKKGPGGAVDSTCLLYDCPDL-----NALPELKHAYD 476
QY 345 PHENIVENKAGVFNKALLSLSVLSEKGLTSPIN-PYKDHMSVLAKYAMNEPNSDEL 403
D 477 PEHPAQPPPARPAEAKPEASASS--ARDGSPSPQPLKUNLSLTKAYVALNQPNSEEL 534
QY 404 LKTSIANGVLPQEFYKEMFEQAKTYQYSNR---SPLEKTSKPLADNSPPTKYDLSLPR 459
D 535 SKIADSVNLPDLYVKFKFEKQACQIRPGQSEPEPSPETGTVINIPAKDEQOPQADGSEPO 594
QY 460 SPVKPMOSITSPSIAELHNSVTSCDPLRLTKSGHFNINIKAVDLHDSRSNT--PSPLN 517
D 595 E-----DSAGQS-----PDKMTSS--PVLPGVSAINGRSCTSPSPSPLN 633

```

```

QY 518 SSTSSKNSSHSYTPNPSSEELQAEPLDLSLPKCMREPKGIATKNTKATSIINDHS 577
D 634 SASRSIQY---FCVADGAQOEPPQVEPLDSLIPKQGE-----LLERST 674
QY 578 VSSSSNS-----DRLNITFTIKKEFSNNNDKNSNPNVFGMPNPSAKPLTPLPQSA 632
D 675 ISSVYQNSVYQVEEPLNLSVKKERQEDSCVTD--SEPVVAVIPSNAPINITALPTVA 732
QY 633 PEPATMPVPQTSIGLARPYPLDQMSFLPHMAYTYPTGAATFADMOORRYQKQFQSG 692
D 733 QLEPTVALDQNSVPCRLALANKQITILIPQAYATYATVSP--ANQEPVVKYIQNGNGD 791
QY 693 DLDGAQDYMGLDMDTSDSCLSRKIKTYSGMVACDLCXTQKSSLRHRYHTG 752
D 792 ERQDTSEGVSTVEDQONSSTTPPKKTRKTENGMVACDLCIKIFQKSSILRHRYHTG 851
QY 753 KRPHOCQICKKPKKHHLIHSRLHSGEKPYQCDCKGRFSGSHSGSYSGHMHRYSYCKR 812
D 852 KRPHECGICKKPKKHHLIHSRLHSGEKPYQCDCKGRFSGSHSGSYSGHMHRYSYCKR 911
QY 813 EAEEREAREAREKHLGPTLMMNRAYIQSTPQGYSDSBERESMPRDSSEKHEKE 872
D 912 EAEERDGTQOEQEAQ---LEALNHNHVGARASP-SQADSDERESLTREDEDSKKEE 966
QY 873 GEEGYGKLRRDQDEEESESESESESESESESESESESESESESESESESESESESE 930
D 967 EEE-----DKEMELQEEKECGNAQEEEEE--EEEFEMDGAKEAAKTDGAVEN 1015
QY 931 KSDHEEDNME 940
D 1016 GAAPQAGSLE 1025

```

RESULT 6

JC4934
delta-erythralin/E2-box factor 1 - mouse
C/Species: Mus musculus (house mouse)
C/Date: 22-Oct-1996 #sequence_revision 01-Nov-1996 #text_change 05-Nov-1999
C/Accession: JC4934
R/Sekido, R.; Takagi, T.; Okanami, M.; Moribe, H.; Yamamura, M.; Higashii, Y.; Kondoh, H.
Gene 173, 221-232, 1996
A/Title: Organization of the gene encoding transcriptional repressor deltaE1 and cross-
A/Reference number: JC4934; MUID:97082972; PMID:8964504
A/Accession: JC4934
A/Status: preliminary; nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-1117 <SRK>
A/Cross-references: DDBJ:D76432; NID:g1027499; PIDN:BA11177.1; PID:dt011838; PID:g102750
C/Comment: This protein is a transcriptional repressor which binds at the E2-box sequence
omain.
C/Genetics:
A/Gene: deltaE1
C/Keywords: zinc finger
F/150-277/Region: zinc fingers
F/882-964/Region: zinc fingers

Query Match 31.8%; Score 1587.5; DB 2; Length 1117;
Best Local Similarity 39.9%; Pred. No.3.1e-75;
Matches 395; Conservative 142; Mismatches 307; Indels 145; Gaps 32;

```

QY 2 LTQAGNRKFKCTECGAKFYKHLKHLRIHSGEKPYECPCNCKRFSHSGSYSHISK 61
D 211 VTQSGNRKFKCTECGAKFYKHLKHLRIHSGEKPYECPCNCKRFSHSGSYSHISK 270
QY 62 KCTGLISVNGRNRNNTTGGSPN-SVSSSPNNSATITOLRNKLENGKRLSSEOTGLIKT 119
D 271 KCTSLMPVNGRPRSGLTQSCSP-SLSTSPSPPTRPQIRQKIEN-KPL--QEPUSVNOI 326
QY 120 TEPLDFNDYKVLMTATHGFSGSSPFMNGGLGATSPLGVPASQSPMCHL---GVGMEADL 176
D 327 TEPLDFY-EPKPIVVASGINCSTPLQNGVFGSGGLQATSSFGGVVAVLTPTVGLVSP 385
QY 177 LGPTMNSNLSEVQKVLQI-VDNTVSRQMDCKTEDISKLGKYMKPCSQPEQGVTS 235

```

```

Db      386 -----SINLSDIQVNLKVAADGNVIRQLTENO-----ASLASKQEQAVASASPIQQGCHSV 436
Qy      236 NIPVGLPVVSHNGATKSIIDYTLKVEAKACLOSLTDSRQISNIKKEKLRLLIDV 295
Db      437 -ISAISLPLVODDTTKIILINSLQPSOLOVVPOLKKEIPAPINSCSEKLPEDLVK 495
Qy      296 TD-DKMIEN-HSISTPFCQCFKESPPG---PIPLHQRVYLCKMNEIKAVLQPHENIV 350
Db      496 SETDSFGARDSTCLCEDC-----FGDLNALPELKHYDPECPAQPPAPATEPRESS 551
Qy      351 PNKAGVFPDNKALLISVLSSEKGLTSPIN-PYKDHMSVUKAYYAMNPNSEDLKISIA 409
Db      552 ASSAP-----NGDLSPPSQPLKNLILKAYYALNAQPSITELSKIDS 595
Qy      410 VGLPOEFYKEMFEOKRVQYNSRSPS-----LERTSKPLAPNSPTTKDSL 457
Db      596 VNLPLDGVKAFKFEKQAGQIP-GQSPDPSPGCTGVNIPTKTDQOPADGNEPOEDSTR 654
Qy      458 PRSPVKNPDSITSPSIAELHNSVTSCDPLRLTKSSHFTNIKAVDKLDHSRNTSPSLNL 517
Db      655 GQSPYKINSSPLP-VGSAMNGSRSC-----TSSPSPLNL 688
Qy      518 STSSKNSHSSSYTNSFSSEELQAPLDLSLPKQMRPEKGIATKNTKATSNLDHNS 577
Db      689 CSARPPQGS---CVAEQAQEEPOVEPLDSLPKQGE-----LLERST 729
Qy      578 VSSSENS-----DEPLNLTFIKKEPNSNNLDNNSNPNVFGMNPFAKELTPLPQSA 632
Db      730 VSSVYQNSVVSQVEPLNLSCKKEPQKDSVTD--SEPVVAVVPPSAPINIAIPLVYA 787
Qy      633 PEPATFMPVQTSISGLRPYPGLDQMSFLPHMAYVPTGATFADMOQR--KYORKQGF 690
Db      788 QLPITVALADQNSVCLALANKOTILLPQVAYTY---SATVSAVGEPPKVIQPNGN 844
Qy      691 QGLDLGAQDYWSGLDMDTSDSCLSRKIKTKTESGMACDLCTFKQSSSLRHKTEH 750
Db      845 QDERDSTSEGVSTVEDQNDSDSTPPKKTREKNGMAACDLCDKIFQKSSILLRHKTEH 904
Qy      751 TGRPHQOQICKKAKKHHLLEHSLRSGEKPYQDCGKRFSGSYSQMNNRYSYC 810
Db      905 TGRHHEGQICKKAKKHHLLEHRLHSGEKPYQDCGKRFSGSYSQMNNRYSYC 964
Qy      811 KRAEEREAEREAREKKGHPTEL--LMNRAVLSITPQGSDEERESMPR--DGSE 866
Db      965 KRGAEDRRAMEGE-----DAGPEVLPVYLATHVGAARASPSQADDERESLREDEDESE 1019
Qy      867 KEHEKEGEE---GYGK-LRRRDGDEEEEEESEENKSMYDPTTIDDEETGDH--- 917
Db      1020 KEEEBEDKEMELQCGKECENPQGESEEEEEEVEEVEADEAHEBAAKTDGT 1079
Qy      918 -----SMDSDSEDKMETKSDHEE 936
Db      1080 VEVGAQAQAGSLQKASSESESESESEQ 1108

RESULT 7
S33641
homeotic protein zfh-1 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 24-Sep-1999
C:Accession: S33641; S27816
R:Portint, M.E.; Lai, Z.; Rubin, G.M.
Mech. Dev. 34, 113-122, 1991
A:Title: The Drosophila zfh-1 and zfh-2 genes encode novel proteins containing both zinc
A:Reference number: S33641; MUID:92001539; PMID:1680376
A:Accession: S33641
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1060 <FOR>
A:Cross-references: EMBL:M63449; NID:g158820; PIDN:AAA23050.1; PID:g158821
C:Genetics:
A:Gene: zfh-1

```

```

A:Cross-references: Flybase:FBgn0004606
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation; zinc finger
P:706-762/Domain: homeobox homology <Hox>

Query Match      16.4%; Score 819.5; DB 2; Length 1060;
Best Local Similarity 29.2%; Pred. No. 4e-35;
Matches 256; Conservative 100; Mismatches 268; Indels 253; Gaps 33;

Qy      9 RRFKTEGCKAPKYVGHILKEHLRHSGEKPYECPNCKKRFSGSYSSHSKCGICGIS 68
Db      359 RRFKTEGCKAPKYVGHILKEHLRHSGEKPYECPNCKKRFSGSYSSHSKCGICGIS 415
Qy      69 VNGMRNN-----IKTSSPNVSYSPTNSAITQLANKLEN-----GKPLSNSEQTGLKXI 119
Db      416 MGLKLNRRALLKRLKESPGSASASRRSPDHGKGLPEQPSLGLPHNPS----- 467
Qy      120 KTEPLDFNDYKVLATHGPGSGSSPPMNGGLGATSP-LGVHPSAQS PMOHLGVMEAPLIG 178
Db      468 -----YFASDAQVQCGSAPAPPPFPHPY-----NMAALIA 499
Qy      179 FP-----TMSNLS-----EVQKVIQIVDNTVSRQKMOCKTEDISKGLYHKDKCSQPE 228
Db      500 FPHNFMMAAGLDPRVHYSIORLLQ--SAAGQOQREEREEOQKQOH--DEETPD 554
Qy      229 EGVTS PNI PVGLPVVSHNGATKSIIDYTLKVEAKACLOSLTDSRQISNIKKEK 288
Db      555 EPKVMDEEBETKEMALTPATEAATPIKRESEA-----SPPEYTRSSQAIKQEG 609
Qy      289 RTLIDLVTDKMIENHS--ISTPSCQCFKESPPGIPIQHQRVYLCKMNEIKAVLQ 345
Db      610 EPLVAERQTPVEEHAHAEVSHADRCGRCSQFHNFPTELQVHEKVLGL--IKELEIQ 666
Qy      346 HENIVPNKAGV-----DNKALLSVLSEK-----LSPINPYKDHMSVKA 390
Db      667 H--FOOQATSPALASASEDEDEDEKMDVEEERPROGSEGRKYRVTAI--EEQOOLQ 722
Qy      391 YYANMNEPNDSEILKISIAVGLPOEFYKEMFE-----ORKVQYNSRS----- 434
Db      723 HYSLNARSRBEFRILARQLDPRVQVFNQNNRERKQSPQNNQAAGAAMPPIDS 782
Qy      435 -PSLERTSKPLAPNSNPTTKDLSLPRSPVVKPMDISITPSIAELHNSVTSCDPLRLTKSS 493
Db      783 QASLTRDQPL-----DLVGRDPLTPKSESPYIA-----PP-----SG 818
Qy      494 HFTNIKAYD-KLDHSRNTSPPLNLSSTSSKNSHSSSTTPNSFSSEELQAEF----- 544
Db      819 EALNPEALNLSRKFTSASMSPASISPSAALYFGAAPPSPSPSOLDSTPFSQAFPG 878
Qy      545 -----LDSLSPKQ-----REPKGIATKNTKATSIINL-DHNSVSSSENSDEPLNLTPI 594
Db      879 LPPTVLPMSLPMALFKPR-PEGDFAS-NHALMSIKLPDTRGTSLSGGSE----- 928
Qy      595 KKEFSNNSNLDNKSNNPVFGMNPFAKPLVPLPQSAFPPTATFMPVQTSIPGLRPYPG 654
Db      929 KRSWRDD-----SRSHDEDFGAGVLMRP----- 953
Qy      655 LDQMSFLPHMAYVPTGATPADMQRRKYORKQFQGLDLDAQDYWSGLDMDTSDSC 714
Db      954 -----KPRRGKVEYTHHAD-----PDLP----- 972
Qy      715 LSRKKIKTKESGMVACDLCTFKQSSSLRHKTEHGTGRPHQOICKKAFGKHHLTEH 774
Db      973 -----YCDQCKRAFAKQSSILAKHKYHSQRPQICEKRAFGKHHLTEH 1019
Qy      775 SRLHSGEKPYQDCGKRFSGSYSQMNNRYSYCK 811
Db      1020 KRLHSGEKPYQSCKLRKRFSGSYSQMNNRYSYCK 1056

RESULT 8
T29204
hypothetical protein F28P9.1 - Caenorhabditis elegans

```


Db 708 HAGEELVYCEECGKAFNRSSNLT-----IHKFIHGE---KPYKCECGKAFNWSSTLTGH 760
 QY 534 S--FSSEELQAEPLDLSLPKQRE-PKGIATKNTKATSIINDHNSVSSSENDPELN 590
 Db 761 KRIHIREK-----PFKCECGKAFIWSSTLTIRKRIH-----TGEKPYK 799
 QY 591 LTFIKFERNSSNNLDKSNINNVFGNPPSAKPLVLPPOSAPPATMPPVQTSIPGLR 650
 Db 800 CEECGKAFSSRSTLT-TKHHTIHTGKPYKCEK-----GKAFKISSALAKKHIIHAGEK 852
 QY 651 PY-----PGIDQWFLPHMATYPTGAATPAD-----MOQRKTKQKQKQ 692
 Db 853 LYKCECGKAFNQSNTLTTHKIIHTKPKSCECDKAFIWSSTLTENHRIHTREKPKYC 912
 QY 693 DILDADQDYMAGL-----DDMTDSCLSRKKIKTESGMVACDCLD 734
 Db 913 EECGKAFQPSHLTTHKMHTEGKPYKCECGKAFSOSTLTTHKIIHTGKPYKCECG 972
 QY 735 KTFQKSSLLRHKYHTGKRPKQCOICKAFKHKHLLIHSRLHSGEKPYQCDKGRPS 794
 Db 973 KAFKRSSTLTTHKIIHTGKPYKCECGKAFSOSTLTFTHTMHTGKPYKCECGKAFN 1032
 QY 795 HSGTSQMHNRHYCYCKEAEEREAERAREKHLGPTLLMRAVLOSTIPQGSDBE 854
 Db 1033 RSSKLTTHKIIHTGKPYKCE-----CGKAFISSSTLNGHKRIH 1072
 QY 855 ERESMPRDESEK-----EHEK--EGEEGY 877
 Db 1073 TREKPYKCECGKAFSOSTLTIRKRIHTGKPY 1106

RESULT 10

500647
 finger protein - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C>Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 31-Dec-1993
 C:Accession: S00647
 R:Ruiz J. 6, Perry-O'Keefe, H.; Melton, D.A.
 EMBL J. 6, 3065-3070, 1987
 A>Title: Xfin: an embryonic gene encoding a multifingered protein in Xenopus.
 A:Reference number: S00647; MUID:88082679; PMID:2826129
 A:Accession: S00647
 A:Molecule type: mRNA
 A:Residues: 1-1350 <RUI>
 A:Cross-references: EMBL:X06021
 A>Note: it is uncertain whether Met-1, Met-11, Met-16, Met-38, or Met-39 is the initiator
 C:Genetics:
 A:Gene: Fin
 C:Keywords: DNA binding; zinc finger

Query Match 7.2%; Score 358; DB 2; Length 1350;
 Best local similarity 19.3%; Pred. No. 7.4e-11;
 Matches 189; Conservative 135; Mismatches 338; Indels 318; Gaps 39;
 QY 6 AGNRKFKTECCEKAFKYLKHLKEHLRIHSGEKPYECPPCKKRFSSGYSYSHISSKCIIG 65
 Db 526 SGEKKFQCAECKCKGTQKSDLVKHIRVHTGKPEKFCILCKKSFQNSDLHKTM----- 578
 QY 66 LISVNGRMNRNITKGTSSSPNSVSSSPTNSAIIOTLRNKLNGKPLMSSEOTGLIKITPELD 125
 Db 579 -----RIHTEKEFPCTYCK-----SPTERSALIKHRTHTG 611
 QY 126 FNDYKVLMTATGSSSPFMNGGLGATSPILGVHPASQSPMOLGVGMARAPLLGPTTMSN 185
 Db 612 ERPHKCSYQCKF-----IQXSALTGHSTHTG-EKP---YPCGCG 649
 QY 186 LSEVQKVLQIVDNTVSRQK-----DCK---TEDISKLGKYHMD-----PCS 225
 Db 650 KSFIGN-----SDLVKHQRIHTGKPYCHTECNKRFTGSSLVK--HRTSGEKPYRCP 702
 QY 226 QPEEGVATSPNIPVGLPVSHNG-----ATKSIIDYTEKVNKA-----C----- 268
 Db 703 QCEKTFIOSSDLVK--HLVYVHNGENPPAATFAHEILIRENTLRSEBDPYPCTECGVVF 759

QY 269 -----LOSITDSRRQISNIKKEKRLTLDLVTDKMIENHSISTPSCQFCCKESF 319
 Db 760 HQPALLKHLATHTTKRYPCNECDKSPFOI-SDLV---KILRHTGGRPHCEPCNGKF 815
 QY 320 PGPILPHOHER-----YLCRMBE-----IKAVLOPHENIVPNKA-----GVFVNDK 361
 Db 816 IQNSDLVGHQRTHTGERPYTSCQCDKFIQRSALTGHRTHTGKPYKCEQCKQCFIONS 875
 QY 362 ALLL-----SSVSEKGLTSPINPYK-----DHMSVLKA 390
 Db 876 DLVHQRHTGKPYHCPDCKRFTGSSLLIKHQRHRIKPYPCGVCKGKFSOSSNLLK 935
 QY 391 YYAMNMEPNSDELLKISIAVGLPOEFVKEWQKRVQYQSNRSRPSLERTSKPLAPNSNP 450
 Db 936 HLKCHSENP-----PALSSELGFVAE-----TQTHP 963
 QY 451 TTQDSL--RSPVPMDS-----ITSPSIAELHNSVTSQDPLRLTKSSH 494
 Db 964 DPVHHIVGDTASYISPEAAGERSFKNDGKCFAHRSVLIKHVRHTGERPYKSCQCTR 1023
 QY 495 FTNIKAVDKLHGSNTPSPNLSTSKNSHSSSYTPNFSSELOEPLDLSLPKQMR 554
 Db 1024 -SFIQKSDLVKHVYTHT-----GERPYKCGLCERSFVEKS-----ALSRQR 1064
 QY 555 EPKGIATKNTKATSIINDHNSVSSSENDPELNLT-----FIKFEPSNNLDKSN 610
 Db 1065 -----VHKNSPVLNSAMEQOQTYNGESKDQNSLVPLQHLVKEE--EAPHVYV---- 1112
 QY 611 PVFGNPPSAKPLVLP--POSAPPATMPPVQTSIPGLRPGLQDMSFLPHMATYTP 669
 Db 1113 -----AVPSLILQSYFP-----ILHP 1130
 QY 670 TGAATFADMQRRKRYQKQGF--OGDLDGADYMSGLDMDTSDCLSRKKIK-KTSSG 726
 Db 1131 KGTFRYSCEGKCFTHRSVFLKIMRMHTTGQPTCKEKGKSFQSSALVGHVRHTGEEK 1190
 QY 727 MYACDLCDKTFQKSSLLRHKYHTGKRPKQCOICKAFKHKHLLIHSRLHSGEKPYQ 786
 Db 1191 PYACSTCKSTIQKSDLAKHQRHTGKPEPYCTYCGKFFIDRSSVVGKSRHTGGERPKC 1250
 QY 787 DKCGRRFSHSGSYSOHNN-----HRSYCKREAEEREAERAREKHLGPTLLMNR 839
 Db 1251 NECTKGFVQKSDLVKHMRTHTGKPYGNCDDRSFSTSAIVRHQR-----MCN- 1299
 QY 840 AYLOSITPQGSYDSEERBSM 859
 Db 1300 -----TGRPYQDEYENSL 1313

RESULT 11

541705
 EVI1 protein - human
 C:Species: Homo sapiens (man)
 C>Date: 25-Dec-1994 #sequence_revision 10-Nov-1995 #text_change 17-Mar-1999
 C:Accession: S41705
 R:Mitani, K.; Ogawa, S.; Tanaka, T.; Miyoshi, H.; Kurokawa, M.; Mano, H.; Yazaki, Y.; Oh,
 EMBL J. 13, 504-510, 1994
 A>Title: Generation of the AML1 - EVI-1 fusion gene in the t(3;21)(q26;q22) causes blast
 A:Reference number: S41704; MUID:94147997; PMID:8313895
 A:Accession: S41705
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1042 <MIT>

Query Match 7.1%; Score 356.5; DB 2; Length 1042;
 Best local similarity 19.5%; Pred. No. 6.2e-11;
 Matches 205; Conservative 144; Mismatches 402; Indels 301; Gaps 38;
 QY 9 RKFKTECGKAFKYLKHLKEHLRIHSGEKPYECNCKKRFSSGYSYSHISSKCIIGLS 68
 Db 101 REYKDCQPKAFNWKSNLIRQMSHDSGKHYESNCACAVFTDPSNLRHRSQHV----- 155

```
QY 69 VNGRMNNIKTSSSPNSPTNSAITQLRNKLENGKPL-----SMSEOTGLIKIRT 121
D 156 -----GARAHACEPCGKTFATSSGLKQHKHHSVKRFCEVCHKSYTQFSNLCRHR 208
QY 122 EPLD-----FNDYKYLMAHGF-SGSSFPNNGIGATSPLVGHSAGSPMOH 167
D 209 MHADCRTOIKCDGCGMFSTSSLNRRRCCKNHPAAGF----- 250
QY 168 LGVGMERAPLGFPTMNSNLSEVQKVLQIVDNTVSRQMKDCTEDISKGYHMKDCPSQ 227
D 251 FGGGISLP--GTPAMDK-----TSMVNMHANGGLADYGC-----ANR 286
QY 228 EEOGVTSPNIP-----PVGL-----PVSHNGATSIIDYTLKYNKACIQSLT 273
D 287 HPAGLFTPTAPGFSFSPGLFPGLYHRPPLIPASSPVKGL--SSTEQTNKSSQ 338
QY 274 TDSRROIINIKKEKATLIDLVTDDKMIENHSISTFSCQFCSEFPGRPLPHOHRYL 333
D 339 -----PLMTHPQILP 348
QY 334 KMMBEIKAVLQPHENIVPNKAGVFDNKALLSSVLSSEKGLTSPINPYKDMSVLKA 393
D 349 ATODILKA--LSKHPISVDNK-----PVELQPERSSEREPFEK 384
QY 394 MNMEPNSEDLKISIAVGLPOEFVKEMFEQRKQYQYNSRSPSLERTSKPLAENSPTTK 453
D 385 ISQSSSDLDVSTPSGDLFTT-----SGSDLESDIDSKKEFKENGKMF 432
QY 454 DSLPPSPVKPM-----SITSPSIAP--LHNSVYSCDPLRLTSSHTNIKAYD 502
D 433 DKVSPLONLASINNKKEYSNHSIFSPSLEQTAVSGAVNDISIKAIASIAEKYFGSGLVG 492
QY 503 KLDHRSNTPSPNLSTSSKNSSSYTPNPSSELOAEPLD-----SLPKOM 553
D 493 LQDKKYGALPYR--SMPLPFPFAFSQSMYR--FPDDRLSLPLKMPQSBGVKYLQKS 549
QY 554 REBKGIATKNTKATKATISINDHNSVSSSENSEDEPLNTFIKKEFSNNDLNKSNPV 613
D 550 SESEFPLTTRKDEKELTPVPSKRPVTATSQDQPLDLSGSRASGTLTTEPRKNHVF 609
QY 614 G-----MNPSPAKPLTYTPPLPQSAFPATMPV-----QTSIP-----GLRPY 652
D 610 GKKGKSNVESRPAISDGL-----QHARPTPEFMDPIYRVEKRLTDPLEALKEXYLPS 663
QY 653 PGL-----DQMSFLPHMAVYTPGAATFADMQQRKRYQKQFGODLLDGAQDVMSGLDMDT 709
D 664 PGLFLHPQMSAIENMA-----EKLSEFSA--LKEPASELLQSVPSMFR 703
QY 710 D---SDSCLSRKKIKTESGMVACDLCDKTFQKSSSILRHKYHTGRPHQCOICCKAKF 766
D 704 NFRAPPAALPEENLIRKQKE--RYTCRYCGKITPPRSANLTRILRTHTGBOYRCKYCDRSFS 762
QY 767 HKHHLIENSR--LHSGEPYQCDCKGRFHSQSYSQM--NHRYSYCKREAEEREAAREA 824
D 763 ISSNLQRHVNINHNKEKPFKCHLCDRCFGQOTNLDRLLKCHENGNNMSGTSSPHSELES 822
QY 825 -----REKHLGP--TELLANRAYLOS-----ITPOGYS--DSE 854
D 823 TGAILDDKEDAYTEIRNFIGNSNHQSOSPRANVERANGSHFKDEKALVYSQNSDLDDE 882
QY 855 ERESMPDGESEKEHEKEGEGYQKL--RRRDGDEEEEBEESBENKSMJDPETIRDEE 912
D 883 EVEDDEVLLDBEDBDNDITGKTGKEPVTSNLHGNPEDDYETSALWMSCTSPRYKERE 942
QY 913 -ETGDHSM-----DSSEDKMETKSDHED 937
D 943 YKSGLSALDHIRHFTDLSLKNRKMEDNQYSBAE 974
```

RESULT 12

A60191
oncogene Evi-1 - human
C:Species: Homo sapiens (man)

C>Date: 20-Feb-1993 #sequence_revision 17-Apr-1993 #text_change 05-Nov-1999
C/Accession: A60191
R/Morishita, K.; Parganas, E.; Douglass, E.C.; Ihle, J.N.
Oncogene 5, 963-971, 1990
A>Title: Unique expression of the human Evi-1 gene in an endometrial carcinoma cell line:
A/Reference number: A60191; MUID:90326419; PMID:2115646
A/Accession: A60191
A/Molecule type: mRNA
A/Residues: 1-1051 <MOR>
A/Cross-references: GB:X54989; NID:950873; PIDN:CAA38735.1; PID:950874
A/Note: the authors translated the codon CCT for residue 85 as Leu
C/Genetics:
A/Genes: GDB:EV11
A/Cross-references: GDB:119889; OMIN:165215
A/Map position: 3q26-3q26
C/Keywords: alternative splicing; DNA binding; zinc finger

Query Match 7.1%; Score 355.5; DB 2; Length 1051;

Best Local Similarity 19.4%; Pred.No.7.1e-11; Matches 199; Conservative 140; Mismatches 395; Indels 293; Gaps 36;

```
QY 9 RKRKTECGKAPKYKHLKELRHSGEKPYECPCNCKKFSHSGSYSHSSKKGICGLIS 68
D 101 REYKDCQCKARFWKSNLIRHQMSHDSGKHTECENCAKVFDPSPNIORHIRSGHV----- 155
QY 69 VNGRMNNIKTSSSPNSPTNSAITQLRNKLENGKPL-----SMSEOTGLIKIRT 121
D 156 -----GARAHACEPCGKTFATSSGLKQHKHHSVKRFCEVCHKSYTQFSNLCRHR 208
QY 122 EPLD-----FNDYKYLMAHGF-SGSSFPNNGIGATSPLVGHSAGSPMOH 167
D 209 MHADCRTOIKCDGCGMFSTSSLNRRRCCKNHPAAGF----- 250
QY 168 LGVGMERAPLGFPTMNSNLSEVQKVLQIVDNTVSRQMKDCTEDISKGYHMKDCPSQ 227
D 251 FGGGISLP--GTPAMDK-----TSMVNMHANGGLADYGC-----ANR 286
QY 228 EEOGVTSPNIP-----PVGL-----PVSHNGATSIIDYTLKYNKACIQSLT 273
D 287 HPAGLFTPTAPGFSFSPGLFPGLYHRPPLIPASSPVKGL--SSTEQTNKSSQ 338
QY 274 TDSRROIINIKKEKATLIDLVTDDKMIENHSISTFSCQFCSEFPGRPLPHOHRYL 333
D 339 -----PLMTHPQILP 348
QY 334 KMMBEIKAVLQPHENIVPNKAGVFDNKALLSSVLSSEKGLTSPINPYKDMSVLKA 393
D 349 ATODILKA--LSKHPISVDNK-----PVELQPERSSEREPFEK 384
QY 394 MNMEPNSEDLKISIAVGLPOEFVKEMFEQRKQYQYNSRSPSLERTSKPLAENSPTTK 453
D 385 ISQSSSDLDVSTPSGDLFTT-----SGSDLESDIDSKKEFKENGKMF 432
QY 454 DSLPPSPVKPM-----SITSPSIAP--LHNSVYSCDPLRLTSSHTNIKAYD 502
D 433 DKVSPLONLASINNKKEYSNHSIFSPSLEQTAVSGAVNDISIKAIASIAEKYFGSGLVG 492
QY 503 KLDHRSNTPSPNLSTSSKNSSSYTPNPSSELOAEPLD-----SLPKOM 553
D 493 LQDKKYGALPYR--SMPLPFPFAFSQSMYR--FPDDRLSLPLKMPQSBGVKYLQKS 549
QY 554 REBKGIATKNTKATKATISINDHNSVSSSENSEDEPLNTFIKKEFSNNDLNKSNPV 613
D 550 SESEFPLTTRKDEKELTPVPSKRPVTATSQDQPLDLSGSRASGTLTTEPRKNHVF 609
QY 614 G-----MNPSPAKPLTYTPPLPQSAFPATMPV-----QTSIP-----GLRPY 652
D 610 GKKGKSNVESRPAISDGL-----QHARPTPEFMDPIYRVEKRLTDPLEALKEXYLPS 663
QY 653 PGL-----DQMSFLPHMAVYTPGAATFADMQQRKRYQKQFGODLLDGAQDVMSGLDMDT 710
D 664 PG-----FLHPQFQPLDQRTMNSAIENMA--EKLSEFSA--LKEPASELLQSVPSMFR 715
```

QY 711 -SDCLSRKKIKKTESGMVACDLCKTFQKSSSLRHKTEHTGKRP HQCICKAPKHKH 769
 Db 716 APPNALPBNLLRKQKE-RYTCRYCGKIFPFRSANLTRHLRTHGEOPYRCKYCDRSEFSISS 774
 QY 770 HLIHSR-LHSGEKRYQCDKCGKRFSHSGSYQHNNHRYCYKRAEAEEREAERAREK 828
 Db 775 NLQRRVRNHNKPKFKCHLCYRCFGQQTNLDRHL-----KKHENG 815
 QY 829 HLGPELMLNRAVYQSIPTFGYSDSEERSEMPRDESEKEHEKE---GEEGYGLRRDG 885
 Db 816 NMSGI-----ATSSPHSELESTGAILDDKEDAVFTIRNFIGNSHSGSPRRV 864
 QY 866 DE-----EEEEEESEENKSM---DTPPTIRIDEETFGDMSDDSEDEKMETKSDH 934
 Db 865 EERNMGSHFEKKALVPSQNSDLDDDEVEDEVLLDER---DEDYDITGKTGKEPVTSNL 921
 QY 935 BEDNMD 941
 Db 922 HEGNPD 928

RESULT 13

A11591

transcription regulator Evi-1 - mouse

C:Species: Mus musculus (house mouse)

C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 02-Jul-1998

C:Accession: A11591

R:Morishita, K.; Parker, D.S.; Mucenaki, M.L.; Jenkins, N.A.; Copeland, N.G.; Ihle, J.N.
 Cell 54, 831-840, 1988

A:Title: Retroviral activation of a novel gene encoding a zinc finger protein in IL-3-de

A:Reference number: A11591; MUID:88311086; PMID:2842066

A:Accession: A11591

A:Molecule type: mRNA

A:Residues: 1-1042 <MOR>

A:Cross-references: GB:M21829

C:Keywords: DNA binding; zinc finger

Query Match 6.9%; Score 345; DB 2; Length 1042;
 Best Local Similarity 19.9%; Pred. No. 2.5e-10;
 Matches 212; Conservative 140; Mismatches 367; Indels 344; Gaps 47;

QY 9 RKFCTEGCGAKFYKHLKHLRIHSGEKRYECPNCKKRFSSHSGSYSHISSKCTIGLS 68
 Db 101 REYKDCQCFKAFNMKSNLIRHQMSHDSGKHYECENCAKYFTDPSNLQRIRISQHV----- 155
 QY 69 VNGRRNNIKTCSSPNVSSTPTNSAIIQLRNKLENGKRL-----SMSBQTGLIKXT 121
 Db 156 -----GARAHACPCGCKTPATISSGLKQKHIISSVKPFICEVCHKSYTQFSNLCRHKR 208
 QY 122 EPLD-----FNDYKVLMTATHGFSG-----SSPFM----- 145
 Db 209 MHADCRTOIKKDCQCMFSTTSLNKHRRFCGKKNHFAAGCFPGQGISLPGFPADKISM 268
 QY 146 -----NGCL-----GATSP-----GVHPSAQSPOHLGVGMEAPLIGF 179
 Db 269 VMMSHANPELADYFGTNRHPAGLTFTAPGFSFSGPLPPS-----GLYHRPPLIPA 320
 QY 160 PTMNSNLSPVQKVLQVDTVSRQKMDCKTEIDISLKGVMMDPCSQPREQGVTSNIRP 239
 Db 321 SEPVGKLGSTEOSNKCQSPLLTHPOLLPATODILKLSR-----PP 362
 QY 240 VG-----LPVSHNGATKSIIDYTLKAYNE-AKACLOSITLDSRQISNKKKELMT 290
 Db 363 VQDNKPVELLP-----ERSSEERLEKISDPOSSESDLDVSTPGSGSLFTSSGDLSS 415
 QY 291 LIDLVTDDKMIENHISITPFSQCFKESPPGPIPLHQRERYLCQKNEEIKAVLQPHENT- 349
 Db 416 --DLSSDKK-----CKEN-----GKMFCKDYKSPLONA 442
 QY 350 -VPNKAGVVDNKALLSLVSEKGLTSPIN-PYKDHMSVLKAYYAMNNEPRSDCLKIS 407
 Db 443 STINKKE-HNNHSVFSASVEQSAVSGAVNDISIKAIASIAEKYF-----GSTGLVGIQ 494

QY 408 I-----AVGLPOEFYKWEF--QRKYQYNSRSRSLERTSKPLADNSNPTTKDLSLPPSP 461
 Db 495 DKVYALPYPMFPLPFPAPFSQSWYP-----PDRDIRSLPL-----KMEQSP 539
 QY 462 VKPMDSITSPSIAELHNSVTSCDPLRLTKSSHFNINIAVDKLSRSRNTSPNLSTSS 521
 Db 540 -----SEVKLQKGSSESPPDLTTTKRD 562
 QY 522 SKNSHSSYTPNSFSESELOAEPLDLSLPKQREBPCKGIATKNKTKATSINDHNSVSS 581
 Db 563 EKPLTSGSKSGTATTS-QDQPLDLSGSKR-----ASGKTLFERNHNVFGEK 613
 QY 562 SENS-D-EPLNLTFTIKCESNSNLDNKNPNPVGNPPF---SAPLYPLPQSAFPAPAT 637
 Db 614 GSNMDTRP-----SSDQSLQHAPTPPF-MDPIYRVEKRLTLPLEALKE----- 657
 QY 638 FMPPVQTSIPGLRPYGL---DQMSFLPHMAYTYTGATPADMQQRKYQKQSGQDL 694
 Db 658 -----KYLRPSPGLFHPQMSAIENMA-----EKLESFSA-L 688
 QY 695 LDGAQDYMSGLDMDTD---SDCLSRKKIKKTESGMVACDLCKTFQKSSSLRHKYEHT 751
 Db 689 KPEASELLQSVPSMSPFAPRNTLPENLIRKQKE-RYTCRYCGKIFPFRSANLTRHLRHT 747
 QY 732 GKRPHQCOICCKAFKHKHLIHSR-LHSGEKRYQCDKCGKRFSHSGSYQHNNHRYCYC 810
 Db 748 GEQPYRCYCDRSPFSSNLRQHVANNHKEKPFCHLCDCRCFGQQTNLDRHL----- 800
 QY 811 KRAEERAEAREAKHGLPTELLMRAVYQSI-----TPQY-----SDSEER 856
 Db 801 -----KKHEINMGSTATSPHSELESAGAILDDKEDAVFTIRNFIGNSHG 848
 QY 857 ESMPRDGESEK--EHEKEGEYGLRRRD--GDEEEEEE---EESEENKMDTPE--- 906
 Db 849 SOSPRNMEERNMGSGFKD-KKALATSQNSDLDDDEVEDEVLLDDEDEDNDIPGRKEL 907
 QY 907 -TIRDEETGDHMSDDSEDEKMETKSD-----HEEDNMDQM 943
 Db 908 GVTIRLDEELPE---DYEEAGALEMSCASPVRYVEEDYKSGL 947

RESULT 14

T14757

hypothetical protein DKFZp572C163.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 01-Dec-2000

C:Accession: T14757

R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, August 1999

A:Reference number: Z18181

A:Accession: T14757

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-701 <MAN>

A:Cross-references: EMBL:AL110217

A:Experimental source: adult subthalamic nucleus; clone DKFZp572C163

C:Genetics:

A:Note: DKFZp572C163.1

C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 6.6%; Score 328.5; DB 2; Length 701;
 Best Local Similarity 19.7%; Pred. No. 1e-09;
 Matches 162; Conservative 87; Mismatches 251; Indels 322; Gaps 31;

QY 7 GNRKFKTECGAKFYKHLKHLRIHSGEKRYECPNCKKRFSSHSGSYSHISSKCTIGL 66
 Db 156 GEKPYECIECGKTFKSTHLRAHQRIHSGEKRYECPNCKKRFSSH-----KTHLS----- 204
 QY 67 ISVNGRMNNIKTSSP---NSVSSPTNSAIIQLRNKLENG-KPLMS--EQTGLIKIX 120
 Db 205 -----VHORVHTGKRYPCNDCKGSPFYNSALRAHQRIHSGEKRYECSDECKT----- 252
 QY 121 TEPLDNDYKVLMTATHG-FSGSSPFMNGLGATSPFGLGVHPSAQSPOHLGVGMEAPL--- 176

```

Db      253 -----FAHNSALAHNRHHTGKPYECNECGRSF---AHISVLKAHQRIHTG- EKYECN 303
Qy      177 -----LGPPTMNSLSEYQKVLQIVDNTVSROKMC-----KTEDISKLGKGYHMK-----D 222
Db      304 ECGRSF-TYNSALRAHQRI-----HTGRKPYECSCDECKTFAHNSALKIHQRIHTGEKPY 356
Qy      223 PCSQPEEQGVTSPIPVGLPVVSHNGATKSIID-YLKEVNEAKACLOSLTDSRRQIS 281
Db      357 ECEJECR-----TPAHNSALRAHQNHHTGEKLYECSECGKTF----- 393
Qy      282 NIKKELKRLTIDLVTDOKMIENHSISPPFCQCKSFPPPIPLHQRHRLCGMNEIKR 341
Db      394 -PQKTLSTHRRHHTGK-----PYCSCKGKTFQSKSYLSGHER-----IHT 435
Qy      342 VLQPHENIYVNAKGVFVFNKALLLSVL--SEKGLTSPINPYKHMSVLKAYYAMNEPN 399
Db      436 GEKPYECNCGK--TFYKALIVHQRIHTGK-----PYECNQ----- 472
Qy      400 SDELKISIAVGLPQEFVKEMFQKRYQYNSRSRPSLEKTSKPLAPNSPTTKDILLPR 459
Db      473 -----CGKTFQQR--THLCAHQRIHTGEKPYECNECGKT----- 504
Qy      460 SPYKPMDSITSPSIAELHNSVTSCDPLRLTKSSHFTNITKAVDKLHRSNTSPILNLS 519
Db      505 -----PADNSALRAHNRHHTGEKPYECNDCG-----KTFKTSHLRA-----HLRT 545
Qy      520 TSSKNSHSSSYTNSFSSEELQAEPLDLSLPKOMREPKGIIATKNKTATKATINLDHNSVS 579
Db      546 RSEKPYECSECGKTSER-----SYNSAH 570
Qy      580 SSENSEDEPLNLTFFIKKFSNSNLDKNSNPVFGMNPFSKAPLYPLPQSAFPATPM 639
Db      571 QRVHTGEKPYECNCGKPFANHSTL----- 595
Qy      640 PVPQTSIPGLRPPGLDQMSFLPHMAVTPYTGAAATPADMQORRYQKQSGODLLDGAQ 659
Db      596 -----RVHQRIH----- 602
Qy      700 DYMSGLDMDTSDSCLSRKIKKTESGMVACDLCDKTFQKSSILRHKYEHTEKRPPOCQ 759
Db      603 -----TGKSYECNDCGKTFQSKSHLSAQRIHTGEKPYECN 639
Qy      760 ICKKAPGKHHLIEHSLHSGEKPYQCKGKAPSHSGSYSQH 802
Db      640 ECGKAFQNSTLRVHQRIHTGEKPYECDECGKTFVRKALRVH 682

RESULT 15
S26823
zinc finger protein ZNF43 - human
N:Alternate names: zinc finger protein kox27
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text_change 01-Dec-2000
C:Accession: S26823; I37967; S10416
R:Overlappg, R.; Trowdale, J.
Nucleic Acids Res. 19, 2921-2928, 1991
A:Title: A gene encoding 22 highly related zinc fingers is expressed in lymphoid cell li
A:Reference number: S26823; MUID:91279444; PMID:1711675
A:Accession: S26823
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-803 <LOV>
R:Residues: 1-803 <LOV>
A:Cross-references: EMBL:X59244; NID:g38031; PIDN:CAA41932.1; PID:g38032
New Biol. 2, 363-374, 1990
A:Title: Multiple genes encoding zinc finger domains are expressed in human T cells.
A:Reference number: I37949; MUID:91145339; PMID:2288909
A:Accession: I37967
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 476-531 <THI>
A:Cross-references: EMBL:X52358; NID:g34160; PIDN:CAA36584.1; PID:g930090

```

C:Genetics:
A:Gene: GDB:ZNF43, HTF6
A:Cross-references: GDB:128653
A:Map position: 19p13.1-19p12
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; zinc finger

Query Match 6.6%; Score 327; DB 2; Length 803;
Best local similarity 19.0%; Pred. No. 1.5e-09;
Matches 160; Conservative 86; Mismatches 196; Indels 402; Gaps 31;

```

Qy      7 GNRKPYECGKAFKXKHLKXELRHSGEKPYECNCKKRFSSHSYSKSCIGL 66
Db      276 GEKPYECGKAFKXKHLKXELRHSGEKPYECNCKKRFSSHSYSKSCIGL 326
Qy      67 ISVGNHRNNIKTSSPNSVSSPTNSAITPOLRNKLENGKPLSMSEOTGLIKTEPLDF 126
Db      327 -----KRIHTGKPYECNCKKRFSSHSYSKSCIGL 363
Qy      127 NDYKVLMAHGFSGSSPFMNGGLGATSPGVHPSAOSPMQHLGVCEADPLGFPYNSNL 186
Db      364 --YKTECEGEAFSRS-----SNL 379
Qy      187 SEYQKVLQIVDNTVSROKMDCTEDI-----SKLGKGYHMDPCSQP---EEQGVTSN 236
Db      380 TKHKK-----IHTGEKPYECGKAFKXKHLKXELRHSGEKPYECNCKKRFSSHSYSKSCIGL 430
Qy      237 IPPVGLPVVSHNGATKSIIDYTLKVNKACLOSLTDSRRQISNIKEKRLTIDLVLT 296
Db      431 WPST--LTKHNRH-----HTGEKPYECGKAFKXKHLKXELRHSGEKPYECNCKKRFSSHSYSKSCIGL 464
Qy      297 DDKMIENHSISPPFCQCKSFPPPIPLHQRH-----YLKRNNEIKAVLQPHENI 349
Db      465 THKRI--HTAEKPYECGKAFKXKHLKXELRHSGEKPYECNCKKRFSSHSYSKSCIGL 509
Qy      350 VPKKAVPVDNKAALLLSVLSKGLT-SPINRYK-----DHMSVLKAYYAMNEPN 400
Db      510 -----GKAFKXKHLKXELRHSGEKPYECNCKKRFSSHSYSKSCIGL 553
Qy      401 DELLKISIAVGLPQEFVKEMFQKRYQYNSRSRPSLEKTSKPLAPNSPTTKDILLPR 460
Db      554 -----HTGEKPYECGKAFKXKHLKXELRHSGEKPYECNCKKRFSSHSYSKSCIGL 577
Qy      461 PVPKMDSITSPSIAELHNSVTSCDPLRLTKSSHFTNITKAVDKLHRSNTSPILNLS 520
Db      578 ----- 577
Qy      521 SSKNSHSSSYTNSF--SSRELQAEPLDLSLPKOMREPKGIIATKNKTATKATINLDHNSVS 579
Db      578 -----HKIHTGEKPYECN----- 592
Qy      580 SSENSEDEPLNLTFFIKKFSNSNLDKNSNPVFGMNPFSKAPLYPLPQSAFPATPM 639
Db      593 -----CGKATQSSNLT-----THKKIH----- 611
Qy      640 PVPQTSIPGLRPPGLDQMSFLPHMAVTPYTGAAATPADMQORRYQKQSGODLLDGAQ 659
Db      612 -----GKPYECGKAFKXKHLKXELRHSGEKPYECNCKKRFSSHSYSKSCIGL 653
Qy      691 QGDLADGADYMSGLDMDTSDSCLSRKIKKTESGMVACDLCDKTFQKSSILRHKYEH 750
Db      654 KW-----SSTLTKXKIHHTGEKPYECGKAFKXKHLKXELRHSGEKPYECNCKKRFSSHSYSKSCIGL 694
Qy      751 TGRKPYECGKAFKXKHLKXELRHSGEKPYECNCKKRFSSHSYSKSCIGL 809
Db      695 TGRKPYECGKAFKXKHLKXELRHSGEKPYECNCKKRFSSHSYSKSCIGL 750
Qy      810 CKK 813
Db      751 TKKQ 754

```

Search completed: January 5, 2004, 16:31:44

Thu Feb 5 09:04:52 2004

us-09-964-238-2.rpt

Page 11

Job time : 35 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2004, 16:23:23 ; Search time 17 Seconds

(without alignments)
2611.365 Million cell updates/sec

Title: US-09-964-238-2

Perfect score: 4991
Sequence: 1 MLTQAGNRKFKTECGRAF.....DKMKETKSDHEDNMEDMG 944

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	ID	Description
1	4891	100.0	1215	1	SIP1_MOUSE	Q9t0g7 mus musculus
2	4869	97.6	1214	1	SIP1_HUMAN	O60315 homo sapien
3	1665	33.4	1114	1	TCF8_CHICK	P36197 gallus gall
4	1657.5	33.2	1124	1	TCF8_HUMAN	P37275 homo sapien
5	1599	31.0	1043	1	TCF8_MESAT	O60542 mesocricetu
6	1587.5	31.8	1117	1	TCF8_MOUSE	O64318 mus musculu
7	1550.5	31.1	1109	1	TCF8_RAT	O62947 rattus norv
8	819.5	16.4	1060	1	ZFHL_DROME	P28166 drosophila
9	407	8.2	1845	1	Z236_HUMAN	O9u136 homo sapien
10	379	7.6	1191	1	EV11_HUMAN	O05481 homo sapien
11	355.5	7.1	1051	1	EV11_HUMAN	O03112 homo sapien
12	354.5	7.1	1350	1	XFIN_XENLA	P08045 xenopus lae
13	345	6.9	1042	1	EV11_MOUSE	P14404 mus musculu
14	341.5	6.8	1167	1	Z208_HUMAN	O43345 homo sapien
15	339.5	6.8	783	1	ZF25_HUMAN	O9u115 homo sapien
16	336	6.7	1276	1	PRD3_HUMAN	O9h822 homo sapien
17	331.5	6.6	751	1	Z184_HUMAN	O99676 homo sapien
18	327	6.6	803	1	ZN43_HUMAN	P17038 homo sapien
19	315.5	6.3	469	1	Z115_HUMAN	P52742 homo sapien
20	310.5	6.2	616	1	ZN61_HUMAN	P35789 homo sapien
21	297	6.0	675	1	ZG20_XENLA	P18714 xenopus lae
22	296	5.9	580	1	ZF35_MOUSE	P15620 mus musculu
23	293	5.9	818	1	KR18_HUMAN	O9hcg1 homo sapien
24	292	5.9	738	1	ZN84_HUMAN	P51523 homo sapien
25	290	5.8	1300	1	SA13_HUMAN	O9bxa9 homo sapien
26	289.5	5.8	898	1	Z071_XENLA	O14587 xenopus lae
27	289.5	5.8	1323	1	SA13_MOUSE	O62255 mus musculu
28	287.5	5.8	947	1	Z268_HUMAN	O14587 homo sapien
29	287.5	5.8	1053	1	SA14_HUMAN	O9u1q4 homo sapien
30	285.5	5.7	636	1	ZF90_MOUSE	O61367 mus musculu
31	285	5.7	856	1	PRD1_MOUSE	O60636 mus musculu
32	281	5.6	626	1	Z189_HUMAN	O75820 homo sapien
33	280.5	5.6	511	1	EGR1_BRARE	P26632 brachydanio

34	280.5	5.6	1324	1	SA11_HUMAN	O9nsc2 homo sapien
35	278.5	5.6	698	1	Z234_HUMAN	O14588 homo sapien
36	278.5	5.6	839	1	Z347_HUMAN	O96687 homo sapien
37	277.5	5.6	780	1	Z084_XENLA	P18753 xenopus lae
38	277	5.5	594	1	ZF37_MOUSE	P17141 mus musculu
39	276	5.5	1029	1	Z197_HUMAN	O14709 homo sapien
40	275	5.5	810	1	Z33A_HUMAN	O06730 homo sapien
41	274.5	5.5	508	1	EGR1_RAT	P08154 rattus norv
42	274.5	5.5	789	1	PRD1_HUMAN	O75626 homo sapien
43	274	5.5	533	1	EGR1_MOUSE	P08046 mus musculu
44	273.5	5.5	521	1	Z286_HUMAN	O9hbt8 homo sapien
45	273.5	5.5	1322	1	SA11_MOUSE	O9et74 mus musculu

ALIGNMENTS

RESULT 1
ID SIP1_MOUSE STANDARD; PRT; 1215 AA.
AC Q9R0G7;
DT 16-OCT-2001 (Rel. 40, Last Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Zinc finger homeobox protein 1b (Smad interacting protein 1).
GN ZFXH1B OR ZFX1B OR SIP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99329065; PubMed=10400677;
RA Vercheren K., Remacle J.E., Collart C., Kraft H., Baker B.S.,
RA Tyranowski P., Nelles U., Muytens G., Su M.-T., Bodmer R.,
RA Smith J.C., Huyebroek D.,
RT "SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad
proteins and binds to 5'-CACCT sequences in candidate target genes";
RT J. Biol. Chem. 274:20498-20498(1999).
RL
CC - FUNCTION: TRANSCRIPTIONAL INHIBITOR THAT BINDS TO DNA SEQUENCE 5'-
CACCT-3' IN DIFFERENT PROMOTERS.
CC - SUBUNIT: BINDS ACTIVATED SMAD1, ACTIVATED SMAD2 AND ACTIVATED
SMAD3; BINDING WITH SMAD4 IS NOT DETECTED.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - SIMILARITY: BELONGS TO DELTA-BF1/ZFH-1 FAMILY OF TWO-HANDED ZINC
FINGER/HOMEOBOXIN PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: AF031116; AAD56590.1; -
DR TRANSFAC: T04864; Zfx1b.
DR MGD: MGI:1344407; Zfx1b.
DR GO: GO:0005634; C:nucleus; IC.
DR GO: GO:0019208; F:phosphatase regulator activity; ISS.
DR GO: GO:0016564; F:transcriptional repressor activity; NAS.
DR GO: GO:0016481; F:negative regulation of transcription; IC.
DR GO: GO:0007399; P:neurogenesis; ISS.
DR GO: GO:0007183; P:SMAD protein heteromerization; ISS.
DR InterPro: IPR001356; Homeobox.
DR InterPro: IPR007087; Znf C2H2.
DR Pfam: PF00096; Zf-C2H2; 8.
DR Prodom: PD000003; Znf C2H2; 1.
DR SMART: SM00389; HOX; 1.
DR SMART: SM00355; Znf C2H2; 8.
DR SMART: PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 6.
KW Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce>
CC or send an email to license@isb-sib.ch).

DR	EMBL; AB011141;	BAA25495.1;	-
DR	EMBL; AY029472;	AALX52081.1;	-
DR	EMBL; AB065607;	BAB40819.1;	-
DR	EMBL; AB015341;	BAA34798.1;	-
DR	TRANSFAC; T05057;	-	-
DR	Genew; H0NC;	14861;	ZFX1B.
DR	MIM; 605802;	-	-
DR	GO; GO:0005634;	C:nucleus;	IC.
DR	GO; GO:0019208;	F:phosphatase regulator activity;	NAS.
DR	GO; GO:0046332;	F:SMAD binding activity;	NAS.
DR	GO; GO:0016564;	F:transcriptional repressor activity;	ISS.
DR	GO; GO:0016481;	P:negative regulation of transcription;	IC.
DR	GO; GO:0007399;	P:neurogenesis;	NAS.
DR	InterPro; IPR001356;	Homeobox.	
DR	InterPro; IPR007087;	Znf C2H2.	
DR	Pfam; PF00096;	zC-ZH2; 8.	
DR	Prodrom; PD000003;	Znf C2H2; 1.	
DR	SMART; SM00355;	ZNF_C2H2; 8.	
DR	PROSITE; PS00028;	ZINC_FINGER_C2H2_1; 5.	
DR	PROSITE; PS50157;	ZINC_FINGER_C2H2_2; 6.	
KV	Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;		
KV	Homeobox; Repressor; Metal-binding; Repeat; Hicshprpn disease.		
FT	DOMAIN	437	487
FT	ZN_FING	211	263
FT	ZN_FING	241	363
FT	ZN_FING	282	304
FT	ZN_FING	310	334
FT	DNA_BIND	644	703
FT	ZN_FING	999	1021
FT	ZN_FING	1027	1049
FT	ZN_FING	1055	1076
FT	DOMAIN	1084	1214
SO	SEQUENCE	1214 AA;	136446 MM; B578BD91335C3FPD CRC64;

Query Match	97.6%	Score 4869	DB 1;	Length 1214;
Best Local Similarity	97.5%	Pred. No. 1,2e-223;		
Matches 919; Conservative	13;	Mismatches 11;	Indels 0;	Gaps 0

Qy	1	MLTQAGNRKFCCTCGCAFKYKHHLKEHLRHHSEKPEECNCKKRFHSHSSYSHHSS	60
Db	272	MLTQAGNRKFCCTCGCAFKYKHHLKEHLRHHSEKPEECNCKKRFHSHSSYSHHSS	331
Qy	61	KKCIGLISVNGMRNNIKTGSSPNSVSSSPITSATITOLNKLKNGKPLSMSEQTGLKIK	120
Db	332	KKCIGLISVNGMRNNIKTGSSPNSVSSSPITSATITOLNKLKNGKPLSMSEQTGLKIK	391
Qy	121	TEPLDPNDYKYLMAHGHFSGSSPPFNNGGIGATSPICGYHSAOSPNOHLGVGHEAPLGGP	180
Db	392	TEPLDPNDYKYLMAHGHFSGTSPFNNGGIGATSPICGVHSAOSPNOHLGVGHEAPLGGP	451
Qy	181	TNNSNLSEVOKLQIVDNTVSRQKMDCKEDISKLKGYHMKPCSQSPQEGYATSPNIPV	240
Db	452	TNNSNLSEVOKLQIVDNTVSRQKMDCKAEELSKLKGYHMKPCSQSPQEGYATSPNIPV	511
Qy	241	GLPVVSHNGATSIIDYTLLEKVENAKACQSLITDTSRROIISNIKEKRLTLLIDLVTDDDM	300
Db	512	GLPVVSHNGATSIIDYTLLEKVENAKACQSLITDTSRROIISNIKEKRLTLLIDLVTDDDM	571
Qy	301	IENHSISTPSCQFCCKESPGRPIPLHÖHERYLCKNNEIKAVLQPHENIVPNKAGVFDN	360
Db	572	IENHSISTPSCQFCCKESPGRPIPLHÖHERYLCKNNEIKAVLQPHENIVPNKAGVFDN	631
Qy	361	KALLISSVLSEGLTSPINPYKDHMSVLKAYYANMNEPNSDELLKISIAVGLPQEFVKEM	420
Db	632	KALLISSVLSEKGMSPINPYKDHMSVLKAYYANMNEPNSDELLKISIAVGLPQEFVKEM	691
Qy	421	FEQRVYQYSNRSRSLERTSKPLAPNSNPTTKDILLPSPVTKPMDSITSPSIAELHNSV	480
Db	691	FEQRVYQYSNRSRSLERTSKPLAPNSNPTTKDILLPSPVTKPMDSITSPSIAELHNSV	751

Dd	692	FEQRKVQYNSRSPLESSKPLAPNSNPPTKOSLIPRSVVKEMDSTISPSIAELNHSV	751
Qy	481	TSOCBPRLTMSHFTNTIKAVDKIDHRSNRPSPLNLSSTSKSHSSSYTPNRSSEL	540
Dd	752	TNCPPLKLTIPSHFTNTIKPEKIDHSSSNTPSPLNLSSTSKSHSSSYTPNRSSEL	811
Qy	541	QAEPLDLSPKQMEPKGIITKTKTKATSINLHNVSSSSENSDEPLNTFIKGFEN	600
Dd	812	QAEPLDLSPKQMEPKSIITKTKTKATSISLHNVSSSSENSDEPLNTFIKGFEN	871
Qy	601	SNINLHNVSNNPFGMNPFSAPLYTLPLPPOGAFPPATMPVQYISIPGLRPYGLDQMSF	660
Dd	872	SNINLHNVSNNPFGMNPFSAPLYTLALPPOGAFPPATMPVQYISIPGLRPYGLDQMSF	931
Qy	661	LPHAAVYPTGAAAFADMQQRKXQKQKQFQGDLLDGAQYMSGLDMDTSDSCISRKXI	720
Dd	932	LPHAAVYPTGAAAFADMQQRKXQKQKQFQGDLLDGAQYMSGLDMDTSDSCISRKXI	991
Qy	721	KTESGMVACDLCKTFOKSSSLRHKYEHTGKRPHOCOTCKAFAKXGHLIEHSLRHS	780
Dd	992	KTESGMVACDLCKTFOKSSSLRHKYEHTGKRPHOCOTCKAFAKXGHLIEHSLRHS	1051
Qy	781	EKPYQCDKCGRFBHSGSYSGQNMHRYSYCKREAREEARAREAREKHLPTELLMRA	840
Dd	1052	EKPYQCDKCGRFBHSGSYSGQNMHRYSYCKREAREEARAREAREKHLPTELLMRA	1111
Qy	841	YLOSITPOGYSDEERESBMPRDSSEKHEKEGEBGYGLRRDGDSEEEEEEESEENKS	900
Dd	1112	YLOSITPOGYSDEERESBMPRDSSEKHEKEGEBGYGLRRDGDSEEEEEEESEENKS	1171
Qy	901	MDTDPETIRDEEETGDSHMDSSDEGKMETYSDHEDNMEGGM	943
Dd	1172	MDTDPETIRDEEETGDSHMDSSDEGKMETYSDHEDNMEGGM	1214

RESULT 3

ID	TCF8 CHICK	STANDARD;	PRT;	1114 AA
----	------------	-----------	------	---------

DT 01-JUN-1994 (Rel. 29, Created)

DT 15-SEP-2003 (Rel. 42, Last annotation update

DE (Delta EF1).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NC Archaeopteryx; Aves; Neornithes; Galliformes; Phasianidae; Phasianinae

OC Galus.
OY NCPT MaxTD-0031

	SENTENCE FROM N A
RN	[1]
BP	

RC	TISSUE=Lens;
PX	MEDLINE-94116444. PubMed-7904558-

RA Funahashi J.-I., Sekido R., Murai K., Kamachi Y., Kondoh H.;
RT "Delta-crystallin enhancer binding protein delta EBF1 is a zinc

RT finger-homedomain protein implicated in postgastrulation embryogenesis¹:

RL	Development	119:433-446(1993)
RN	[2]	

RP SEQUENCE FROM N.A
RC TISSUE=Embryo:

RX MEDLINE=97082972; PubMed=8964504;
RA Sekido R., Takagi T., Okanami M., Moribe H., Yamamura M., Higashi Y.,

RT "Organization of the gene encoding transcriptional repressor deltaEFl
K4 London H.;

RL Gene 173:227-232 (1996) .
 RI and cross-species conservation of its domains. . . ;

CC LENS-SPECIFIC TRANSCRIPTION. IT BINDS AS WELL MANY OTHER NON-LENS

CC SPECIFIC DNA SEQUENCES.
CC
CC -|- SUBUNIT: MAY INTERACT WITH CTBP1 (BY SIMILARITY).
CC

CC - TISSUE SPECIFICITY: EXPRESSION IS DEVELOPMENTALLY REGULATED WITH

RL Science 254:1791-1794(1991).

CC -1- FUNCTION: INHIBITS INTERLEUKIN-2 (IL-2) GENE EXPRESSION. MAY BE
 CC RESPONSIBLE FOR TRANSCRIPTIONAL REPRESSION OF THE IL-2 GENE.
 CC ENHANCES OR REPASSES THE PROMOTER ACTIVITY OF THE ATPIAL GENE
 CC DEPENDING ON THE QUANTITY OF CDNA AND ON THE CELL TYPE.

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- TISSUE SPECIFICITY: EXPRESSED IN HEART AND SKELETAL MUSCLE, BUT
 CC NOT IN LIVER, SPLEEN, OR PANCREAS.

CC -1- SIMILARITY: BELONGS TO DELTA-BF1/ZFH-1 FAMILY OF TWO-HANDED ZINC
 CC FINGER/HOMEOBOX MAIN PROTEINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL, D15050; BAA03646.1; .

DR EMBL, U12170; AAA20602.1; .

DR EMBL, M81699; .; NOT_ANNOTATED_CDS.

DR PIR, JX0293; JX0293.

DR TRANSFAC, T00625; .

DR Genew, HGNC:11642; TCF8.

DR MIM, 189909; .

DR GO, GO:0003713; F:transcription co-activator activity; TAS.

DR GO, GO:0003714; F:transcription co-repressor activity; TAS.

DR GO, GO:0003700; F:transcription factor activity; TAS.

DR GO, GO:0008270; F:zinc ion binding activity; TAS.

DR GO, GO:0008283; P:cell proliferation; TAS.

DR GO, GO:0006955; P:immune response; TAS.

DR GO, GO:0001122; P:negative regulation of transcription from P. .; TAS.

DR InterPro, IPR01356; Homeobox.

DR InterPro, IPR007087; Znf C2H2.

DR Pfam, PF00966; Zf-C2H2; 7.

DR Prodom, PD00010; Homeobox; 1.

DR Prodom, PD000003; Znf C2H2; 2.

DR SMART, SM00389; HOX; 1.

DR SMART, SM00355; ZNF C2H2; 7.

DR PROSITE, PS00028; ZINC_FINGER_C2H2_1; 5.

DR PROSITE, PS0157; ZINC_FINGER_C2H2_2; 6.

KW Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
 KW Homeobox; Repressor; Activator; Metal-binding; Repeat.

FT ZN_FING 170 193 C2H2-TYPE.

FT ZN_FING 200 222 C2H2-TYPE.

FT ZN_FING 240 262 C2H2-TYPE.

FT ZN_FING 268 292 C2H2-TYPE (ATYPICAL).

FT DNA_BIND 581 640 HOMEBOX-LIKE.

FT ZN_FING 904 926 C2H2-TYPE.

FT ZN_FING 932 954 C2H2-TYPE.

FT ZN_FING 960 981 C2H2-TYPE (ATYPICAL).

FT DOMAIN 989 1124 GLU-RICH (ACIDIC).

FT CONFLICT 420 420 V -> I (IN REF. 2).

FT CONFLICT 609 609 E -> Q (IN REF. 3).

FT CONFLICT 654 654 I -> T (IN REF. 2).

FT CONFLICT 672 672 L -> H (IN REF. 3).

FT CONFLICT 681 681 D -> S (IN REF. 3).

SO SEQUENCE 1124 AA; 124073 MW; 0A2714CC37C84BD1 CRC64;

Query Match 33.2%; Score 1657.5; DB 1; Length 1124;
 Best Local Similarity 41.9%; Pred. No. 1.7e-71;
 Matches 406; Conservative 147; Mismatches 306; Indels 107; Gaps 32;

QY 2 LTGAGNRKFKTEGCKAFKTKHLEHRIHSGEYPCNPKKRFHSHSGSYSHHSK 61
 DB 231 VTQSCNKKFKCTEGCKAFKTKHLEHRIHSGEYPCNPKKRFHSHSGSYSHHSK 290

QY KCIIGLISVNGRNNIKTG--SSPNSVSSPTNSAITOLRNLKNGKPLMSSEOTGLKI 119
 DB 231 KCIISLIPNGRRTGLKTSQCSPP-SLSASPSGSPRPRQIKRIEN-KPI--OEQLSVQOI 346

QY 120 KTEPLDFNDYKVLATHTGFGSGSPFMNGGLGATSPLGVAHPSAQSPMQL--GVGMEAPL 176

DB 347 KTEPVY-EFKPIVVASGINSCTPLONGVFTGGPLOTSSPQGNVAVLPTVGLVSP 405
 QY LGPFTMNSNLEEVQVLOI-VDNVTSRQMOCKTEDIKLGHYMKDCSPQEEGVTS 235
 DB 406 -----SINLSDIQNLVAVDGNVIOVLE---NQNANLASKQETINASPIQGGHVS 456
 QY 236 NIPVGLPVVSHNGATSKSIIDTYLEKVAACLOSLTDSRRQISNKKERLRLD 295
 DB 457 -ISALSLPVDQDGTTLIIIVSLQPSQLOVVPQNLKKNPVATNSCKSEKLPEDLTVK 515
 QY 296 TD-DKMTFNH-SISTPSCQFCXSPGPI-PLHCHERYLCKMNEIRAVLQPHENIYPN 352
 DB 516 SEKKSFGGVNDSTCLLDCDC---PEDINALBELKAY-----DLKQPTOP-----PP 560
 QY 353 KAGVFDKALLLSVSEKGLTSPINRYKDMSTLKYVYAMNNEPNSDELKISIAVGL 412
 DB 561 LPAEAERPESSVSATDGNLSPQPLKULSLKLYALNAPSAEELSKIDSVNL 620
 QY 413 POEFVKEWFEQRYOXSNSRPSLERTSKPLAPNSPTTQDSLPRSPVXPMDSITSPS 472
 DB 621 PLDYVKKWFEKMGQAGQIS-----VQSEBSPERPKV-----NIPAKNNDQPGAN 666
 QY 473 IAEILHNSVTSQDPLRLTKSSHFTNRYKAVDKLDHSRSTPSPPLNISSTSSKXSSSTYP 532
 DB 667 ANEPDQSTVNLQSPDKMTNS---PVLPGSTTNGSRSTSPSPPLNLSSTSHNTQGYLYTA 723
 QY 533 NSFSEELQAPRLDLSLPKMRREPKGIIATNKKATISINLDHNSVSSSSNS-----DE 587
 DB 724 EG-AQEEQVPELDLSLPKQGE-----LLESTTIVYQNSVSVQEE 766
 QY 588 PLNTLFKKESSNSNLDNKNPNPFVGNPPSAKPLTLPPOSAPPATMPPVQTSIP 647
 DB 767 PLNLSCKAKEQKQKSCVTD--SEPVVNVIPPSANPINAIPTVRAQLPTIYALADQNSVP 824
 QY 648 GLRPYGLDQMSFLPHNAVITYPTGAATPADMQQRKRYKQKQFGQDLIDGADIVMSGILD 707
 DB 825 CLRALANKQTLLPQVAVTYSTVSP-AVOEPLKVIQPNQODERODTSEGVSNVED 883
 QY 708 MTDSDSCSRKKIKKTESGAWACLDCTPKQSSSLSLHKYEHGKRRPHOQOICKAKFKH 767
 DB 884 QNDSSTPPKKMKRTENGMAACDLCDKI PQSSSLHKKYEHGKRRPHOQOICKAKFKH 943
 QY 768 KHLHLEHRLHSGEYPCYCDKCGKRFHSHSGSYSGHMNRVYCYCKREAEREAREAREK 827
 DB 944 KHLHLEHRLHSGEYPCYCDKCGKRFHSHSGSYSGHMNRVYCYCKREAEREAREAREK 999
 QY 828 GHLGTELLMRAVLIQITP-QGYSDSERESMPR--DGSEKHEKEKEGEE-----GYGK 879
 DB 1000 -EAGP-ETLSNEHVQARASPSQG--DSDERESLTREEDDESEKEEEDKEWELQEEKE 1055
 QY 880 LRRDGDDEEEERESKNSMDPTIRDEEERGHSMDSDD-----GKNET 930
 DB 1056 CEKPGDEEEEBEVEEBEVEE-BAENBGEAKTEGLMKODRASQASLQGVKESSE 1114
 QY 931 KSDHEEDN 938
 DB 1115 QVSEKTN 1122

RESULT 5
 TCF8_MESAU
 ID_TCF8_MESAU STANDARD; PRT; 1043 AA.
 AC 060542;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Transcription factor 8 (zinc finger protein Bzp).
 GN TCF8 OR BZP.
 OS Mesocricetus auratus (golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.

OX NCBI_TaxID=10036;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Insulinoma;
 RX MEDLINE=95021206; PubMed=7935395;
 RA Franklin A.J., Jettion T.L., Shelton K.D., Magnuson M.A.;
 RT "B2P, a novel serum-responsive zinc finger protein that inhibits gene transcription.";
 RL Mol. Cell. Biol. 14:6773-6788(1994).
 CC -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL REPRESSOR. BINDS TO E-BOX SEQUENCES IN THE IMMNOGLOBULIN HEAVY CHAIN ENHANCER AS WELL AS IN THE REGULATORY REGIONS OF MANY OTHER TISSUE-SPECIFIC GENES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO DELTA-EF1/ZFH-1 FAMILY OF TWO-HANDED ZINC FINGER/HOMEDOMAIN PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L13856; AAC37667.1; -.
 CC PIR: A56037; A56037.
 CC InterPro: IPR001356; Homeobox.
 CC InterPro: IPR007087; Znf_C2H2.
 CC Pfam: PF00996; zf-C2H2; 7.
 CC ProDom: PD000010; Homeobox; 1.
 CC ProDom: PD000003; Znf_C2H2; 2.
 CC SMART: SM00389; HOX; 1.
 CC SMART: SM00355; Znf_C2H2; 7.
 CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
 CC PROSITE: PS50157; ZINC_FINGER_C2H2_2; 6.
 CC DR Transcription regulation; DNA-binding; Nuclear protein; zinc-finger; Homeobox; Repressor; Activator; Metal-binding; Repeat.
 CC KW FT ZN_FING 94 117 C2H2-TYPE.
 CC FT ZN_FING 124 146 C2H2-TYPE.
 CC FT ZN_FING 164 186 C2H2-TYPE.
 CC FT ZN_FING 192 216 C2H2-TYPE (ATYPICAL).
 CC FT DNM_BIND 504 563 HOMEBOX-LIKE.
 CC FT ZN_FING 827 849 C2H2-TYPE.
 CC FT ZN_FING 855 877 C2H2-TYPE.
 CC FT ZN_FING 883 904 C2H2-TYPE (ATYPICAL).
 CC FT DOMAIN 912 1043 GLU-RICH (ACIDIC).
 CC SQ SEQUENCE 1043 AA; 114202 MW; 5F8BE23DD1BC667 CRC64;
 Query Match 32.0%; Score 1599; DB 1; Length 1043;
 Best Local Similarity 40.2%; Pred. No. 9.2e-69;
 Matches 390; Conservative 142; Mismatches 308; Indels 130; Gaps 30;
 QY 2 LNOGANNRKKCECGKAPKFKHLEKHLRIHSGEKYRECPNKKRFSHSGSYSHISK 61
 DB 155 VTQSGNRRKFCCEGAFKFKHLEKHLRIHSGEKYRECPNKKRFSHSGSYSHISK 214
 QY 62 KCIGLISVNGRMNRIKTCGSSPN-SVSSSPTNSAITQLANKLNGKPLMSQGTGLKIK 120
 DB 215 KCIQLMPVNGRPSGLKTCGSPSLASGSPTRQIRKIKIN-KPL-QEPLSNQIK 271
 QY 121 TEPLDNDYKVLMTATGFGSSGSPFNG--GLG---ATSLGVHPSAGSPMCHLVGME 173
 DB 272 TEPLVDY-EFKPIVAVAGINCSTPLQNGVFGGQLQNTSSPGCVAVVLP----VGLV 326
 QY 174 APLLGFTMNSNLSVQKQI-QI-VDNTVSHQKMDCTEDISKLGKYMPCQPREQCV 232
 DB 327 SPT----SINSDIQNLKVALDGNVRQVLENNQASL----ASKQDEAV 368
 QY 233 TSPN----IPVGLPVVSHNGATKSIIDYTLKVNKAKACLSLTDSRQISNIK 284
 DB 369 NASSIOGGSHVSAISLPLVDQDGTITLINSLSQPSQLQVPPQNLKENAPFNSCK 428
 QY 285 KEKLRTLLIDLVTDKMIENHSISTPSCQCKSPGPIPLHGHERYLCMNBEIKAVIQ 344

DB 429 SEKLPE--DLTVKSEKDKFGDGVADSGTCLLYDDCGDI-----NALPELKHVD 476
 QY 345 PHENIVPNKAGYFVNDKALLSSVLEKGLTSPIN-PYDHNHSLKAYAYAMNEPNSDEL 403
 DB 477 PEHPAPPPPAPEAEKEPSAASS--ARDDBLSPSQPLNKLNLKAYAYALNAPNSEEL 534
 QY 404 LKSIAGVLPQEFVKEWEPQKRVYQYSNR---SPSLERTSKPLAPNSPTTKSLPR 459
 DB 535 SKIADSVNPLDVLVKKFKFMQAGIIPQSPSPSPETGTVINPAKSDQPOPADSSEQA 594
 QY 460 SPKPMDSITSSIAELHNSVTSCEPLRLTYSNHTNLIKADKLDHRSNT--PSPLNL 517
 DB 595 E-----DSASGGS-----PLMTSS--PVLPGSALNGSSCTSPSPPLNL 633
 QY 518 SSTSSKNSHSSSTPNSFSEELQAEPLDLSPKQREPKGIATQNKTKATISINDHNS 577
 DB 634 SSKRLQGY--FCVADGQAEQEPQVEPLDLSLPKQGE-----LLERST 674
 QY 578 VSSSSSENS-----DEPLNFTIKKESNSNNLDNKSNNPFGKNNPSPAKPLTLPPOA 632
 DB 675 ISSVQNSVYSQVEEPLNLSVCKEPEQDSCTVD--SEPVNVVIPSANPINALIPTVA 732
 QY 633 PPATMPPVQTSIPGLRTPYGLDQMSFLPHAYTTPGATPADMQQRKTKQKQFQ 692
 DB 733 QLPTIVALDQNSVPCLRALANKQTILLPOVAYTATVSP-AMQEPVVKYIQPNQND 791
 QY 693 DILDDGADYMSGLDDTDSDSCLSRKIKITSGMYACDLCKTFQKSSSLRLHKYHTG 752
 DB 792 ERDITSESEGVYEDQNDSDSTPPKKTRTKNGMYACDLCKTIPOKSSSLRLHKYHTG 851
 QY 753 KRPHQCIQKAPKFKHLEHSLHSGEKPYQCDKCGKRFSGSGSYQHMNHRYSYCKR 812
 DB 852 KRPHQCIQKAPKFKHLEHSLHSGEKPYQCDKCGKRFSGSGSYQHMNHRYSYCKR 911
 QY 813 EAEEERAAERAREKHLQPTELMNRATLQSTTPGYSDBERREMPDGESEKHEK 872
 DB 912 EAEEERGTEQOEQEA---LEALMNEHVGARASP-SQADSERESTLTEEDDSKEER 966
 QY 873 GEEGYKLLRRQDEEESESESEKSMQDTPETIRDEETGHDHSDSS--EDGKMET 930
 DB 967 EEE-----DKMEHLOEKECKGAQAEEREE--EEEEEDGAKDAAKTQDAVEN 1015
 QY 931 KSDHEEDNME 940
 DB 1016 GAAPQAGSLE 1025
 RESULT 6
 TCF8 MOUSE
 ID TCF8 MOUSE STANDARD; PRT; 1117 AA.
 AC 064318; 062519;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Transcription factor 8 (Zinc finger homeobox protein 1a) (MEB1) (Delta EF1).
 GN TCF8 OR ZFX1A OR ZFX1A OR ZFX1A.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Embryo;
 RX MEDLINE=97082972; PubMed=8964504;
 RA Sekido R., Takagi T., Moribe H., Yamamura M., Higashi Y., Kondoh H.;
 RT "Organization of the gene encoding transcriptional repressor deltaEF1 and cross-species conservation of its domains.";
 RL Gene 173:227-232(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Spleen;


```

RA  Mu Y., Montoya G.D., Rubin S.E., Brodie S.G., Jenkins N.,
RA  Williams T.M.,
RL  Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN  [3]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Brain;
RX  MEDLINE=96194821; PubMed=8647466;
RA  Genetta T., Kadesch T.,
RT  "Cloning of a cDNA encoding a mouse transcriptional repressor
RT  displaying striking sequence conservation across vertebrates.";
RL  Gene 169:289-290(1996).
CC  -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL REPRESSOR. BINDS TO E-BOX
CC  SEQUENCES IN THE IMMUGLOBULIN HEAVY CHAIN ENHANCER AS WELL AS IN
CC  THE REGULATORY REGIONS OF MANY OTHER TISSUE-SPECIFIC GENES.
CC  -1- SUBCELLULAR LOCATION: Nuclear.
CC  -1- SIMILARITY: BELONGS TO DELTA-BFL/ZFH-1 FAMILY OF TWO-HANDED ZINC
CC  FINGER/HOMEOBOXIN PROTEINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; U26259; AAA67564.1; -.
DR  EMBL; D76432; BA01177.1; -.
DR  EMBL; L48363; AAB08442.1; -.
DR  PIR; JC4934; JC4934.
DR  TRANSFAC; T03915; -.
DR  MGD; MGI:1344313; Zfx1a.
DR  InterPro; IPR001356; Homeobox.
DR  InterPro; IPR007087; Znf C2H2.
DR  Pfam; PF00096; Zf-C2H2; 7.
DR  ProDom; PD000010; Homeobox; 1.
DR  ProDom; PD000003; Znf C2H2; 2.
DR  SMART; SM00389; HOK; 1.
DR  SMART; SM00355; Znf_C2H2; 7.
DR  PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR  PROSITE; PS50157; ZINC_FINGER_C2H2_2; 6.
KM  Transcription regulation: DNA-binding; Nuclear protein; Zinc-finger;
KM  Homeobox; Repressor; Activator; Metal-binding; Repeat.
FT  ZN_FING 150 173 C2H2-TYPE.
FT  ZN_FING 180 202 C2H2-TYPE.
FT  ZN_FING 220 242 C2H2-TYPE.
FT  ZN_FING 248 272 C2H2-TYPE (ATYPICAL).
FT  DNA_BIND 559 618 HOMEBOX-LIKE.
FT  ZN_FING 882 904 C2H2-TYPE.
FT  ZN_FING 910 932 C2H2-TYPE.
FT  ZN_FING 938 959 C2H2-TYPE (ATYPICAL).
FT  DOMAIN 969 1117 GLU-RICH (ACIDIC).
FT  CONFLICT 33 33 S -> A (IN REF. 3).
FT  CONFLICT 57 57 M -> K (IN REF. 3).
FT  CONFLICT 125 125 MISSING (IN REF. 3).
FT  CONFLICT 146 146 F -> L (IN REF. 3).
FT  CONFLICT 313 313 E -> EV (IN REF. 3).
FT  CONFLICT 353 353 G -> A (IN REF. 3).
FT  CONFLICT 400 400 V -> L (IN REF. 3).
FT  CONFLICT 461 461 K -> E (IN REF. 3).
FT  CONFLICT 528 528 K -> KK (IN REF. 3).
FT  CONFLICT 664 664 S -> T (IN REF. 3).
FT  CONFLICT 744 744 E -> L (IN REF. 3).
FT  CONFLICT 822 822 T -> A (IN REF. 3).
FT  CONFLICT 1062 1062 MISSING (IN REF. 3).
SO  SEQUENCE 1117 AA; 122464 MW; D1FAC2D04BD3437 CRC64;

Query Match 31.8%; Score 1587.5; DB 1; Length 1117;
Best Local Similarity 39.9%; Pred. No. 3.5e-68;
Matches 395; Conservative 142; Pident 307; Indels 145; Gaps 32;

```

```

Db      211 VTGSGNRKRKFCSTCGAFKFKHKLKXHLRIHSEKDEYECNCKGRSHSGSYSHHSK 270
Qy      62 KCICLISVNGMRNRNITG--SGSNVSSTPFTSAITQLRNLKNGKPLMSBOTGLIKI 119
Db      271 KCILSMFVNGPRSGLTSSQCSSP-SLSTSGSPTRPQIRQIKIN--KPL--QEBLSVNOI 326
Qy      120 KTEBLDNDYKVLMAHTGFSSGSPFMNNGUGATSPGLVHPAOSPMOHL--GYGMEAPL 176
Db      327 KTEBVDI-EKFPVIVASGINSTYPLQNGVSSGGQLATSSPGGVAVVLPVGLVSP 385
Qy      177 LGFPTMNSNI,SEVOKV,OI--VDNTVSRQMDCKTEDI,SKLGYMKDPCSOPEEGVTS 235
Db      386 -----SINISDIQNLKAVDAGVNIQVLETNQ--ASLASKEQEAVASAP1QCGGSHV 436
Qy      236 NIPVGLPVVSHNGATYSIIDTYLEKVNKAQ,LSQSLTTSRRQISNIKKEKLTLDLV 295
Db      437 -ISATISPLVDQDSTKIIINYSLEOPSQ,OVNPQN,KEI,PAPTNCSSEKLPEDLTVK 495
Qy      296 TD-DKML,EN-HSISTPPSCOPCKSPFG--PILHQHERL,CMQMEI,RAVLDPHENIV 350
Db      496 SETKSPFEGARDSTCLCEDC-----GGLNALRELKHYPECAQPPPPAPATEKPESS 551
Qy      351 PNKAGVFDNKALLLSVLSSEKGLTSPIN--PYCDHMSVLKAYYMMNNEPNSDELLKTSIA 409
Db      552 ASSAG-----NGDLSGQPLKULSLIKAYYALNMQSTREELSKADS 595
Qy      410 VGLPQEFVKWEFEQKRYOYVNSRSPS-----LERTSKPLAPNSNPTTKSL 457
Db      596 VNLPLDGVKFKFEQAGQ,IP-QGSPDPSPGTSVNI,PTKTDQ,POPADNEPOEDSTR 654
Qy      458 PRSPVTKMDISTTSGSI,DELHNSVTS,CPPLRLTSSHFTNIKAVDKDHS,SNTPSPLNT 517
Db      655 GQSPVTKR,SSVLP-VSSAMNGSRSC-----TSSPSPLNT 688
Qy      518 SSTSKSHSSSYTPNFSSEI,LAEP,LDLSL,PKRMEPKGIIATKXKTKATSIN,LDNS 577
Db      689 CSANPPOGYS---CVAGAQDEEPVPELDLSLPQOGE-----LIERST 729
Qy      578 VSSSENS-----DEPLNT,LTIKKE,PFNSNN,LDKSNNPVFGMNP,FSAKPL,LYPL,POSA 632
Db      730 VSSVYQNSVYSVGEPLNT,SCAKKEPQKDS,CVTD--SEPVVNVV,PSANP,INIAI,PTVTA 787
Qy      633 PFPATMP,PQVTS,IPGLR,PI,PGLDQMS,F,PLMATY,PTGAT,PFADMOOR--KYQKQGF 690
Db      788 QLPITVIA,DAQNS,VP,CCRALA,NKQTL,IP,OVATY--SATVSPA,VOEPVK,VIOPGN 844
Qy      691 QGDL,LDGADQMSGLDMDMTD,SD,CLSRKXI,KTESGMYACD,CPKTKOKSS,ILRH,KEH 750
Db      845 QDEQODTSSBEV,ST,VEBQ,NDSD,PPPKKTKRKTENGMYACD,CKITQKSS,SLRH,KEH 904
Qy      751 TGRAPHO,COICCKA,FKFKH,LI,HSR,LSHSGE,KP,QCQKCGKRF,SHSGSYQ,MMHNR,SYC 810
Db      905 TGRKPHRCIGICRKA,FKFKH,LI,EMHRLHSGE,KP,QCQKCGKRF,SHSGSYQ,MMHNR,SYC 964
Qy      811 KGEAEERBAERAREK,HLG,LPTEL--LMNR,AY,QSITPQ,SYDSER,RESMPR--DGESE 866
Db      965 KRGAEDRDAMEQE-----DAGPEVL,PLV,LA,TEH,GA,LA,SP,QA,SD,RES,ITREDE,DESE 1019
Qy      867 KEHKKEEBE-----CYGK-LRRQDDEEBE,EESE,ENK,SDT,DE,ETIR,DEE,ETODH---- 917
Db      1020 KEEBEEDKE,MEEL,OEGB,ECEN,POCE,EEEE,EEEE,EEEE,EEEE,VEADEABE,HA,AKTDGT 1079
Qy      918 -----SMDDSEDKMET,SKD,EE 936
Db      1080 VEVGAQAQAGSLRQKASBSEMESESESEBQ 1108

RESULT 7
ID_TCF8_RAT TCF8_RAT STANDARD; PRT; 1109 AA.
AC 062947: 062948:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

15-SEP-2003 (Ref. 42, Last annotation update)
 DE Transcription factor 8 (zinc finger homeodomain enhancer-binding
 DE protein) (Zfhpf).
 GN TCF8.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE OF 6-1109 FROM N.A., AND ALTERNATIVE SPLICING.
 RA MEDLINE=96365389; PubMed=8769566;
 RX Cabanillas A.M., Darling D.S.;
 RT "Alternative splicing gives rise to two isoforms of Zfhpf, a zinc
 RT finger/homeodomain protein that binds T3-response elements.";
 RL DNA Cell Biol. 15:643-651(1996).
 CC -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL REPRESSOR. BINDS TO E-BOX
 CC SEQUENCES IN THE IMMUGLOBULIN HEAVY CHAIN ENHANCER AS WELL AS IN
 CC THE REGULATORY REGIONS OF MANY OTHER TISSUE-SPECIFIC GENES (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=Zfhpf-1;
 CC IsoId=662947-1; Sequence=Displayed;
 CC Name=2; Synonyms=Zfhpf-2;
 CC IsoId=662947-2; Sequence=VSP_006881;
 CC -1- SIMILARITY: BELONGS TO DELTA-BF1/ZFH-1 FAMILY OF TWO-HANDED ZINC
 CC FINGER/HOMEODOMAIN PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U51583; AAB17130.1; -.
 DR EMBL; U51584; AAB17131.1; -.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR007087; Znf C2H2.
 DR Pfam; PF00096; zf-C2H2; 7.
 DR ProDom; PD000010; Homeobox; 1.
 DR ProDom; PD000003; Znf C2H2; 2.
 DR SMART; SMO0389; HOX; 1.
 DR SMART; SMO0355; Znf C2H2; 7.
 DR PROSITE; PS00028; ZINC FINGER C2H2 1; 5.
 DR PROSITE; PS50157; ZINC FINGER C2H2 2; 6.
 KW Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
 KW Homeobox; Repressor; Activator; Metal-binding; Repeat;
 KM Homeobox; Repressor; Activator; Metal-binding; Repeat;
 FT Alternative splicing.
 FT ZN_FING 150 173 C2H2-TYPE.
 FT ZN_FING 180 202 C2H2-TYPE.
 FT ZN_FING 220 242 C2H2-TYPE.
 FT ZN_FING 248 272 C2H2-TYPE (ATYPICAL).
 FT DNA_BIND 559 618 HOMEOBOX-LIKE.
 FT ZN_FING 881 903 C2H2-TYPE.
 FT ZN_FING 909 931 C2H2-TYPE.
 FT ZN_FING 937 958 C2H2-TYPE (ATYPICAL).
 FT DOMAIN 968 1109 GLU-RICH (ACIDIC).
 FT VASPLIC 1 198 Misling (in isoform 2).
 FT /FTID=VSP_006881.
 SQ SEQUENCE 1109 AA; 121626 MW; BEFE29JC8795DDA6 CRC64;
 Query Match 31.1%; Score 1550.5; DB 1; Length 1109;
 Best Local Similarity 39.2%; Pred. No. 2e-66; Indels 157; Gaps 35;
 Matches 387; Conservative 150; Mismatches 294;
 2 LTGAGNKKFKCTCEGKAFKYGHHLKEHLRIHSGEPYECPCCKRFSHSGSYSHISSK 61
 :|||
 211 VTGSGGNKKFKCTCEGKAFKYGHHLKEHLRIHSGEPYECPCCKRFSHSGSYSHISSK 270
 62 KCIIGLISVNGRRNNIKTG--SSPNVSASSPFLNSAITQLRNKLNGKPLSSEQTGLIKI 119

[illegible]

OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydriidae; Drosophilidae; Drosophila.
 NCBI_Taxid=7227;
 [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=92001539; PubMed=1680376;
 RA Fortini M.E., Lai Z., Rubin G.M.;
 RT "The Drosophila zfh-1 and zfh-2 genes encode novel proteins
 containing both zinc-finger and homeobox motifs.";
 RL Mech. Dev. 34:113-122(1991).
 CC - FUNCTION: INVOLVED IN THE DEVELOPMENT OF THE EMBRYONIC CENTRAL
 CC NERVOUS SYSTEM, EMBRYONIC MESODERM AND ADULT MUSCULATURE.
 CC - SUBCELLULAR LOCATION: Nuclear (Probable).
 CC - TISSUE SPECIFICITY: MESODERM AND MESODERMALLY-DERIVED STRUCTURES
 CC IN THE EMBRYO INCLUDING THE DORSAL VESSEL, SUPPORT CELLS OF THE
 CC GONADS, AND SEGMENT-SPECIFIC ARRAYS OF ADULT MUSCLE PRECURSOR.
 CC ALSO IDENTIFIED IN MOTOR NEURONS OF DEVELOPING CNS.
 CC SIMILARITY: BELONGS TO DELTA-EF1/ZFH-1 FAMILY OF TWO-HANDED ZINC
 CC FINGER/HOMEOBOX PROTEINS.
 CC - SIMILARITY: Contains 9 C2H2-type zinc fingers.
 CC - SIMILARITY: Contains 1 homeobox domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sb.ch/announce/>
 CC or send an email to license@isb-sb.ch).
 CC -----
 DR EMBL: M63449; AAA2050.1; -.
 DR PIR: S33641; S33641.
 DR HSSP: P08153; 1ZED.
 DR TRANSFAC: T00919; -.
 DR FLYBase; FBgn0004606; zfh1.
 DR GO: GO:0007498; P:mesoderm development; IEP.
 DR GO: GO:0007399; P:neurogenesis; IEP.
 DR InterPro: IPR001356; Homeobox.
 DR InterPro: IPR007087; Znf_C2H2.
 DR Pfam: PF00046; homeobox; 1.
 DR Pfam: PF00096; zf-C2H2; 9.
 DR ProDom: PD000010; Homeobox; 1.
 DR ProDom: PD000003; Znf_C2H2; 2.
 DR SMART: SM00389; HOX; 1.
 DR SMART: SM00355; Znf_C2H2; 9.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS00021; HOMEBOX_2; 1.
 DR PROSITE: PS00028; ZINC FINGER C2H2_1; 6.
 DR PROSITE: PS00157; ZINC FINGER C2H2_2; 9.
 KW Zinc-finger; Metal-binding; DNA-binding; Homeobox; Nuclear protein;
 KW Repeat.
 FT ZN_FING 74 97 C2H2-TYPE 1.
 FT DOMAIN 222 252 GLN-RICH (OPA-REPEAT).
 FT ZN_FING 295 317 C2H2-TYPE 2.
 FT ZN_FING 330 352 C2H2-TYPE 3.
 FT ZN_FING 361 383 C2H2-TYPE 4.
 FT ZN_FING 389 413 C2H2-TYPE 5.
 FT ZN_FING 413 435 C2H2-TYPE 6.
 FT ZN_FING 435 457 C2H2-TYPE 7.
 FT ZN_FING 457 479 C2H2-TYPE 8.
 FT ZN_FING 479 501 C2H2-TYPE 9.
 FT ZN_FING 501 523 C2H2-TYPE 10.
 FT ZN_FING 523 545 C2H2-TYPE 11.
 FT ZN_FING 545 567 C2H2-TYPE 12.
 FT ZN_FING 567 589 C2H2-TYPE 13.
 FT ZN_FING 589 611 C2H2-TYPE 14.
 FT ZN_FING 611 633 C2H2-TYPE 15.
 FT ZN_FING 633 655 C2H2-TYPE 16.
 FT ZN_FING 655 677 C2H2-TYPE 17.
 FT ZN_FING 677 699 C2H2-TYPE 18.
 FT ZN_FING 699 721 C2H2-TYPE 19.
 FT ZN_FING 721 743 C2H2-TYPE 20.
 FT ZN_FING 743 765 C2H2-TYPE 21.
 FT ZN_FING 765 787 C2H2-TYPE 22.
 FT ZN_FING 787 809 C2H2-TYPE 23.
 FT ZN_FING 809 831 C2H2-TYPE 24.
 FT ZN_FING 831 853 C2H2-TYPE 25.
 FT ZN_FING 853 875 C2H2-TYPE 26.
 FT ZN_FING 875 897 C2H2-TYPE 27.
 FT ZN_FING 897 919 C2H2-TYPE 28.
 FT ZN_FING 919 941 C2H2-TYPE 29.
 FT ZN_FING 941 963 C2H2-TYPE 30.
 FT ZN_FING 963 985 C2H2-TYPE 31.
 FT ZN_FING 985 1007 C2H2-TYPE 32.
 FT ZN_FING 1007 1029 C2H2-TYPE 33.
 FT ZN_FING 1029 1051 C2H2-TYPE 34.
 FT ZN_FING 1051 1073 C2H2-TYPE 35.
 FT ZN_FING 1073 1095 C2H2-TYPE 36.
 FT ZN_FING 1095 1117 C2H2-TYPE 37.
 FT ZN_FING 1117 1139 C2H2-TYPE 38.
 FT ZN_FING 1139 1161 C2H2-TYPE 39.
 FT ZN_FING 1161 1183 C2H2-TYPE 40.
 FT ZN_FING 1183 1205 C2H2-TYPE 41.
 FT ZN_FING 1205 1227 C2H2-TYPE 42.
 FT ZN_FING 1227 1249 C2H2-TYPE 43.
 FT ZN_FING 1249 1271 C2H2-TYPE 44.
 FT ZN_FING 1271 1293 C2H2-TYPE 45.
 FT ZN_FING 1293 1315 C2H2-TYPE 46.
 FT ZN_FING 1315 1337 C2H2-TYPE 47.
 FT ZN_FING 1337 1359 C2H2-TYPE 48.
 FT ZN_FING 1359 1381 C2H2-TYPE 49.
 FT ZN_FING 1381 1403 C2H2-TYPE 50.
 FT ZN_FING 1403 1425 C2H2-TYPE 51.
 FT ZN_FING 1425 1447 C2H2-TYPE 52.
 FT ZN_FING 1447 1469 C2H2-TYPE 53.
 FT ZN_FING 1469 1491 C2H2-TYPE 54.
 FT ZN_FING 1491 1513 C2H2-TYPE 55.
 FT ZN_FING 1513 1535 C2H2-TYPE 56.
 FT ZN_FING 1535 1557 C2H2-TYPE 57.
 FT ZN_FING 1557 1579 C2H2-TYPE 58.
 FT ZN_FING 1579 1601 C2H2-TYPE 59.
 FT ZN_FING 1601 1623 C2H2-TYPE 60.
 FT ZN_FING 1623 1645 C2H2-TYPE 61.
 FT ZN_FING 1645 1667 C2H2-TYPE 62.
 FT ZN_FING 1667 1689 C2H2-TYPE 63.
 FT ZN_FING 1689 1711 C2H2-TYPE 64.
 FT ZN_FING 1711 1733 C2H2-TYPE 65.
 FT ZN_FING 1733 1755 C2H2-TYPE 66.
 FT ZN_FING 1755 1777 C2H2-TYPE 67.
 FT ZN_FING 1777 1799 C2H2-TYPE 68.
 FT ZN_FING 1799 1821 C2H2-TYPE 69.
 FT ZN_FING 1821 1843 C2H2-TYPE 70.
 FT ZN_FING 1843 1865 C2H2-TYPE 71.
 FT ZN_FING 1865 1887 C2H2-TYPE 72.
 FT ZN_FING 1887 1909 C2H2-TYPE 73.
 FT ZN_FING 1909 1931 C2H2-TYPE 74.
 FT ZN_FING 1931 1953 C2H2-TYPE 75.
 FT ZN_FING 1953 1975 C2H2-TYPE 76.
 FT ZN_FING 1975 1997 C2H2-TYPE 77.
 FT ZN_FING 1997 2019 C2H2-TYPE 78.
 FT ZN_FING 2019 2041 C2H2-TYPE 79.
 FT ZN_FING 2041 2063 C2H2-TYPE 80.
 FT ZN_FING 2063 2085 C2H2-TYPE 81.
 FT ZN_FING 2085 2107 C2H2-TYPE 82.
 FT ZN_FING 2107 2129 C2H2-TYPE 83.
 FT ZN_FING 2129 2151 C2H2-TYPE 84.
 FT ZN_FING 2151 2173 C2H2-TYPE 85.
 FT ZN_FING 2173 2195 C2H2-TYPE 86.
 FT ZN_FING 2195 2217 C2H2-TYPE 87.
 FT ZN_FING 2217 2239 C2H2-TYPE 88.
 FT ZN_FING 2239 2261 C2H2-TYPE 89.
 FT ZN_FING 2261 2283 C2H2-TYPE 90.
 FT ZN_FING 2283 2305 C2H2-TYPE 91.
 FT ZN_FING 2305 2327 C2H2-TYPE 92.
 FT ZN_FING 2327 2349 C2H2-TYPE 93.
 FT ZN_FING 2349 2371 C2H2-TYPE 94.
 FT ZN_FING 2371 2393 C2H2-TYPE 95.
 FT ZN_FING 2393 2415 C2H2-TYPE 96.
 FT ZN_FING 2415 2437 C2H2-TYPE 97.
 FT ZN_FING 2437 2459 C2H2-TYPE 98.
 FT ZN_FING 2459 2481 C2H2-TYPE 99.
 FT ZN_FING 2481 2503 C2H2-TYPE 100.
 FT ZN_FING 2503 2525 C2H2-TYPE 101.
 FT ZN_FING 2525 2547 C2H2-TYPE 102.
 FT ZN_FING 2547 2569 C2H2-TYPE 103.
 FT ZN_FING 2569 2591 C2H2-TYPE 104.
 FT ZN_FING 2591 2613 C2H2-TYPE 105.
 FT ZN_FING 2613 2635 C2H2-TYPE 106.
 FT ZN_FING 2635 2657 C2H2-TYPE 107.
 FT ZN_FING 2657 2679 C2H2-TYPE 108.
 FT ZN_FING 2679 2701 C2H2-TYPE 109.
 FT ZN_FING 2701 2723 C2H2-TYPE 110.
 FT ZN_FING 2723 2745 C2H2-TYPE 111.
 FT ZN_FING 2745 2767 C2H2-TYPE 112.
 FT ZN_FING 2767 2789 C2H2-TYPE 113.
 FT ZN_FING 2789 2811 C2H2-TYPE 114.
 FT ZN_FING 2811 2833 C2H2-TYPE 115.
 FT ZN_FING 2833 2855 C2H2-TYPE 116.
 FT ZN_FING 2855 2877 C2H2-TYPE 117.
 FT ZN_FING 2877 2899 C2H2-TYPE 118.
 FT ZN_FING 2899 2921 C2H2-TYPE 119.
 FT ZN_FING 2921 2943 C2H2-TYPE 120.
 FT ZN_FING 2943 2965 C2H2-TYPE 121.
 FT ZN_FING 2965 2987 C2H2-TYPE 122.
 FT ZN_FING 2987 3009 C2H2-TYPE 123.
 FT ZN_FING 3009 3031 C2H2-TYPE 124.
 FT ZN_FING 3031 3053 C2H2-TYPE 125.
 FT ZN_FING 3053 3075 C2H2-TYPE 126.
 FT ZN_FING 3075 3097 C2H2-TYPE 127.
 FT ZN_FING 3097 3119 C2H2-TYPE 128.
 FT ZN_FING 3119 3141 C2H2-TYPE 129.
 FT ZN_FING 3141 3163 C2H2-TYPE 130.
 FT ZN_FING 3163 3185 C2H2-TYPE 131.
 FT ZN_FING 3185 3207 C2H2-TYPE 132.
 FT ZN_FING 3207 3229 C2H2-TYPE 133.
 FT ZN_FING 3229 3251 C2H2-TYPE 134.
 FT ZN_FING 3251 3273 C2H2-TYPE 135.
 FT ZN_FING 3273 3295 C2H2-TYPE 136.
 FT ZN_FING 3295 3317 C2H2-TYPE 137.
 FT ZN_FING 3317 3339 C2H2-TYPE 138.
 FT ZN_FING 3339 3361 C2H2-TYPE 139.
 FT ZN_FING 3361 3383 C2H2-TYPE 140.
 FT ZN_FING 3383 3405 C2H2-TYPE 141.
 FT ZN_FING 3405 3427 C2H2-TYPE 142.
 FT ZN_FING 3427 3449 C2H2-TYPE 143.
 FT ZN_FING 3449 3471 C2H2-TYPE 144.
 FT ZN_FING 3471 3493 C2H2-TYPE 145.
 FT ZN_FING 3493 3515 C2H2-TYPE 146.
 FT ZN_FING 3515 3537 C2H2-TYPE 147.
 FT ZN_FING 3537 3559 C2H2-TYPE 148.
 FT ZN_FING 3559 3581 C2H2-TYPE 149.
 FT ZN_FING 3581 3603 C2H2-TYPE 150.
 FT ZN_FING 3603 3625 C2H2-TYPE 151.
 FT ZN_FING 3625 3647 C2H2-TYPE 152.
 FT ZN_FING 3647 3669 C2H2-TYPE 153.
 FT ZN_FING 3669 3691 C2H2-TYPE 154.
 FT ZN_FING 3691 3713 C2H2-TYPE 155.
 FT ZN_FING 3713 3735 C2H2-TYPE 156.
 FT ZN_FING 3735 3757 C2H2-TYPE 157.
 FT ZN_FING 3757 3779 C2H2-TYPE 158.
 FT ZN_FING 3779 3801 C2H2-TYPE 159.
 FT ZN_FING 3801 3823 C2H2-TYPE 160.
 FT ZN_FING 3823 3845 C2H2-TYPE 161.
 FT ZN_FING 3845 3867 C2H2-TYPE 162.
 FT ZN_FING 3867 3889 C2H2-TYPE 163.
 FT ZN_FING 3889 3911 C2H2-TYPE 164.
 FT ZN_FING 3911 3933 C2H2-TYPE 165.
 FT ZN_FING 3933 3955 C2H2-TYPE 166.
 FT ZN_FING 3955 3977 C2H2-TYPE 167.
 FT ZN_FING 3977 3999 C2H2-TYPE 168.
 FT ZN_FING 3999 4021 C2H2-TYPE 169.
 FT ZN_FING 4021 4043 C2H2-TYPE 170.
 FT ZN_FING 4043 4065 C2H2-TYPE 171.
 FT ZN_FING 4065 4087 C2H2-TYPE 172.
 FT ZN_FING 4087 4109 C2H2-TYPE 173.
 FT ZN_FING 4109 4131 C2H2-TYPE 174.
 FT ZN_FING 4131 4153 C2H2-TYPE 175.
 FT ZN_FING 4153 4175 C2H2-TYPE 176.
 FT ZN_FING 4175 4197 C2H2-TYPE 177.
 FT ZN_FING 4197 4219 C2H2-TYPE 178.
 FT ZN_FING 4219 4241 C2H2-TYPE 179.
 FT ZN_FING 4241 4263 C2H2-TYPE 180.
 FT ZN_FING 4263 4285 C2H2-TYPE 181.
 FT ZN_FING 4285 4307 C2H2-TYPE 182.
 FT ZN_FING 4307 4329 C2H2-TYPE 183.
 FT ZN_FING 4329 4351 C2H2-TYPE 184.
 FT ZN_FING 4351 4373 C2H2-TYPE 185.
 FT ZN_FING 4373 4395 C2H2-TYPE 186.
 FT ZN_FING 4395 4417 C2H2-TYPE 187.
 FT ZN_FING 4417 4439 C2H2-TYPE 188.
 FT ZN_FING 4439 4461 C2H2-TYPE 189.
 FT ZN_FING 4461 4483 C2H2-TYPE 190.
 FT ZN_FING 4483 4505 C2H2-TYPE 191.
 FT ZN_FING 4505 4527 C2H2-TYPE 192.
 FT ZN_FING 4527 4549 C2H2-TYPE 193.
 FT ZN_FING 4549 4571 C2H2-TYPE 194.
 FT ZN_FING 4571 4593 C2H2-TYPE 195.
 FT ZN_FING 4593 4615 C2H2-TYPE 196.
 FT ZN_FING 4615 4637 C2H2-TYPE 197.
 FT ZN_FING 4637 4659 C2H2-TYPE 198.
 FT ZN_FING 4659 4681 C2H2-TYPE 199.
 FT ZN_FING 4681 4703 C2H2-TYPE 200.
 FT ZN_FING 4703 4725 C2H2-TYPE 201.
 FT ZN_FING 4725 4747 C2H2-TYPE 202.
 FT ZN_FING 4747 4769 C2H2-TYPE 203.
 FT ZN_FING 4769 4791 C2H2-TYPE 204.
 FT ZN_FING 4791 4813 C2H2-TYPE 205.
 FT ZN_FING 4813 4835 C2H2-TYPE 206.
 FT ZN_FING 4835 4857 C2H2-TYPE 207.
 FT ZN_FING 4857 4879 C2H2-TYPE 208.
 FT ZN_FING 4879 4901 C2H2-TYPE 209.
 FT ZN_FING 4901 4923 C2H2-TYPE 210.
 FT ZN_FING 4923 4945 C2H2-TYPE 211.
 FT ZN_FING 4945 4967 C2H2-TYPE 212.
 FT ZN_FING 4967 4989 C2H2-TYPE 213.
 FT ZN_FING 4989 5011 C2H2-TYPE 214.
 FT ZN_FING 5011 5033 C2H2-TYPE 215.
 FT ZN_FING 5033 5055 C2H2-TYPE 216.
 FT ZN_FING 5055 5077 C2H2-TYPE 217.
 FT ZN_FING 5077 5099 C2H2-TYPE 218.
 FT ZN_FING 5099 5121 C2H2-TYPE 219.
 FT ZN_FING 5121 5143 C2H2-TYPE 220.
 FT ZN_FING 5143 5165 C2H2-TYPE 221.
 FT ZN_FING 5165 5187 C2H2-TYPE 222.
 FT ZN_FING 5187 5209 C2H2-TYPE 223.
 FT ZN_FING 5209 5231 C2H2-TYPE 224.
 FT ZN_FING 5231 5253 C2H2-TYPE 225.
 FT ZN_FING 5253 5275 C2H2-TYPE 226.
 FT ZN_FING 5275 5297 C2H2-TYPE 227.
 FT ZN_FING 5297 5319 C2H2-TYPE 228.
 FT ZN_FING 5319 5341 C2H2-TYPE 229.
 FT ZN_FING 5341 5363 C2H2-TYPE 230.
 FT ZN_FING 5363 5385 C2H2-TYPE 231.
 FT ZN_FING 5385 5407 C2H2-TYPE 232.
 FT ZN_FING 5407 5429 C2H2-TYPE 233.
 FT ZN_FING 5429 5451 C2H2-TYPE 234.
 FT ZN_FING 5451 5473 C2H2-TYPE 235.
 FT ZN_FING 5473 5495 C2H2-TYPE 236.
 FT ZN_FING 5495 5517 C2H2-TYPE 237.
 FT ZN_FING 5517 5539 C2H2-TYPE 238.
 FT ZN_FING 5539 5561 C2H2-TYPE 239.
 FT ZN_FING 5561 5583 C2H2-TYPE 240.
 FT ZN_FING 5583 5605 C2H2-TYPE 241.
 FT ZN_FING 5605 5627 C2H2-TYPE 242.
 FT ZN_FING 5627 5649 C2H2-TYPE 243.
 FT ZN_FING 5649 5671 C2H2-TYPE 244.
 FT ZN_FING 5671 5693 C2H2-TYPE 245.
 FT ZN_FING 5693 5715 C2H2-TYPE 246.
 FT ZN_FING 5715 5737 C2H2-TYPE 247.
 FT ZN_FING 5737 5759 C2H2-TYPE 248.
 FT ZN_FING 5759 5781 C2H2-TYPE 249.
 FT ZN_FING 5781 5803 C2H2-TYPE 250.
 FT ZN_FING 5803 5825 C2H2-TYPE 251.
 FT ZN_FING 5825 5847 C2H2-TYPE 252.
 FT ZN_FING 5847 5869 C2H2-TYPE 253.
 FT ZN_FING 5869 5891 C2H2-TYPE 254.
 FT ZN_FING 5891 5913 C2H2-TYPE 255.
 FT ZN_FING 5913 5935 C2H2-TYPE 256.
 FT ZN_FING 5935 5957 C2H2-TYPE 257.
 FT ZN_FING 5957 5979 C2H2-TYPE 258.
 FT ZN_FING 5979 6001 C2H2-TYPE 259.
 FT ZN_FING 6001 6023 C2H2-TYPE 260.
 FT ZN_FING 6023 6045 C2H2-TYPE 261.
 FT ZN_FING 6045 6067 C2H2-TYPE 262.
 FT ZN_FING 6067 6089 C2H2-TYPE 263.
 FT ZN_FING 6089 6111 C2H2-TYPE 264.
 FT ZN_FING 6111 6133 C2H2-TYPE 265.
 FT ZN_FING 6133 6155 C2H2-TYPE 266.
 FT ZN_FING 6155 6177 C2H2-TYPE 267.
 FT ZN_FING 6177 6199 C2H2-TYPE 268.
 FT ZN_FING 6199 6221 C2H2-TYPE 269.
 FT ZN_FING 6221 6243 C2H2-TYPE 270.
 FT ZN_FING 6243 6265 C2H2-TYPE 271.
 FT ZN_FING 6265 6287 C2H2-TYPE 272.
 FT ZN_FING 6287 6309 C2H2-TYPE 273.
 FT ZN_FING 6309 6331 C2H2-TYPE 274.
 FT ZN_FING 6331 6353 C2H2-TYPE 275.
 FT ZN_FING 6353 6375 C2H2-TYPE 276.
 FT ZN_FING 6375 6397 C2H2-TYPE 277.
 FT ZN_FING 6397 6419 C2H2-TYPE 278.
 FT ZN_FING 6419 6441 C2H2-TYPE 279.
 FT ZN_FING 6441 6463 C2H2-TYPE 280.
 FT ZN_FING 6463 6485 C2H2-TYPE 281.
 FT ZN_FING 6485 6507 C2H2-TYPE 282.
 FT ZN_FING 6507 6529 C2H2-TYPE 283.
 FT ZN_FING 6529 6551 C2H2-TYPE 284.
 FT ZN_FING 6551 6573 C2H2-TYPE 285.
 FT ZN_FING 6573 6595 C2H2-TYPE 286.
 FT ZN_FING 6595 6617 C2H2-TYPE 287.
 FT ZN_FING 6617 6639 C2H2-TYPE 288.
 FT ZN_FING 6639 6661 C2H2-TYPE 289.
 FT ZN_FING 6661 6683 C2H2-TYPE 290.
 FT ZN_FING 6683 6705 C2H2-TYPE 291.
 FT ZN_FING 6705 6727 C2H2-TYPE 292.
 FT ZN_FING 6727 6749 C2H2-TYPE 293.
 FT ZN_FING 6749 6771 C2H2-TYPE 294.
 FT ZN_FING 6771 6793 C2H2-TYPE 295.
 FT ZN_FING 6793 6815 C2H2-TYPE 296.
 FT ZN_FING 6815 6837 C2H2-TYPE 297.
 FT ZN_FING 6837 6859 C2H2-TYPE 298.
 FT ZN_FING 6859 6881 C2H2-TYPE 299.
 FT ZN_FING 6881 6903 C2H2-TYPE 300.
 FT ZN_FING 6903 6925 C2H2-TYPE 301.
 FT ZN_FING 6925 6947 C2H2-TYPE 302.
 FT ZN_FING 6947 6969 C2H2-TYPE 303.
 FT ZN_FING 6969 6991 C2H2-TYPE 304.
 FT ZN_FING 6991 7013 C2H2-TYPE 305.
 FT ZN_FING 7013 7035 C2H2-TYPE 306.
 FT ZN_FING 7035 7057 C2H2-TYPE 307.
 FT ZN_FING 7057 7079 C2H2-TYPE 308.
 FT ZN_FING 7079 7101 C2H2-TYPE 309.
 FT ZN_FING 7101 7123 C2H2-TYPE 310.
 FT ZN_FING 7123 7145 C2H2-TYPE 311.
 FT ZN_FING 7145 7167 C2H2-TYPE 312.
 FT ZN_FING 7167 7189 C2H2-TYPE 313.
 FT ZN_FING 7189 7211 C2H2-TYPE 314.
 FT ZN_FING 7211 7233 C2H2-TYPE 315.
 FT ZN_FING 7233 7255 C2H2-TYPE 316.
 FT ZN_FING 7255 7277 C2H2-TYPE 317.
 FT ZN_FING 7277 7299 C2H2-TYPE 318.
 FT ZN_FING 7299 7321 C2H2-TYPE 319.
 FT ZN_FING 7321 7343 C2H2-TYPE 320.
 FT ZN_FING 7343 7365 C2H2-TYPE 321.
 FT ZN_FING 7365 7387 C2H2-TYPE 322.
 FT ZN_FING 7387 7409 C2H2-TYPE 323.
 FT ZN_FING 7409 7431 C2H2-TYPE 324.
 FT ZN_FING 7431 7453 C2H2-TYPE 325.
 FT ZN_FING 7453 7475 C2H2-TYPE 326.
 FT ZN_FING 7475 7497 C2H2-TYPE 327.
 FT ZN_FING 7497 7519 C2H2-TYPE 328.
 FT ZN_FING 7519 7541 C2H2-TYPE 329.
 FT ZN_FING 7541 7563 C2H2-TYPE 330.
 FT ZN_FING 7563 7585 C2H2-TYPE 331.
 FT ZN_FING 7585 7607 C2H2-TYPE 332.
 FT ZN_FING 7607 7629 C2H2-TYPE 333.
 FT ZN_FING 7629 7651 C2H2-TYPE 334.
 FT ZN_FING 7651 7673 C2H2-TYPE 335.
 FT ZN_FING 7673 7695 C2H2-TYPE 336.
 FT ZN_FING 7695 7717 C2H2-TYPE 337.
 FT ZN_FING 7717 7739 C2H2-TYPE 338.
 FT ZN_FING 7739 7761 C2H2-TYPE 339.
 FT ZN_FING 7761 7783 C2H2-TYPE 340.
 FT ZN_FING 7783 7805 C2H2-TYPE 341.
 FT ZN_FING 7805 7827 C2H2-TYPE 342.
 FT ZN_FING 7827 7849 C2H2-TYPE 343.
 FT ZN_FING 7849 7871 C2H2-TYPE 344.
 FT ZN_FING 7871 7893 C2H2-TYPE 345.
 FT ZN_FING 7893 7915 C2H2-TYPE 346.
 FT ZN_FING 7915 7937 C2H2-TYPE 347.
 FT ZN_FING 7937 7959 C2H2-TYPE 348.
 FT ZN_FING 7959 7981 C2H2-TYPE 349.
 FT ZN_FING 7981 8003 C2H2-TYPE 350.
 FT ZN_FING 8003 8025 C2H2-TYPE 351.
 FT ZN_FING 8025 8047 C2H2-TYPE 352.
 FT ZN_FING 8047 8069 C2H2-TYPE 353.
 FT ZN_FING 8069 8091 C2H2-TYPE 354.
 FT ZN_FING 8091 8113 C2H2-TYPE 355.
 FT ZN_FING 8113 8135 C2H2-TYPE 356.
 FT ZN_FING 8135 8157 C2H2-TYPE 357.
 FT ZN_FING 8157 8179 C2H2-TYPE 358.
 FT ZN_FING 8179 8201 C2H2-TYPE 359.
 FT ZN_FING 8201 8223 C2H2-TYPE 360.
 FT ZN_FING 8223 8245 C2H2-TYPE 361.
 FT ZN_FING 8245 8267 C2H2-TYPE 362.
 FT ZN_FING 8267 8289 C2H2-TYPE 363.
 FT ZN_FING 8289 8311 C2H2-TYPE 364.
 FT ZN_FING 8311 8333 C2H2-TYPE 365.
 FT ZN_FING 8333 8355 C2H2-TYPE 366.
 FT ZN_FING 8355 8377 C2H2-TYPE 367.
 FT ZN_FING 8377 8399 C2H2-TYPE 368.
 FT ZN_FING 8399 8421 C2H2-TYPE 369.
 FT ZN_FING 8421 8443 C2H2-TYPE 370.
 FT ZN_FING 8443 8465 C2H2-TYPE 371.
 FT ZN_FING 8465 8487 C2H2-TYPE 372.
 FT ZN_FING 8487 8509 C2H2-TYPE 373.
 FT ZN_FING 8509 853

CC TISSUE=Kidney; PubMed=10458916;
 CC MEDLINE=93389731; Mason R.M.;
 CC Holmes D.I.; Mahab N.A.;
 CC "Cloning and characterization of ZNF236, a glucose-regulated Kruppel-
 CC like zinc-finger gene mapping to human chromosome 18q22-q23.";
 CC Genomics 60:105-109(1999). AS A TRANSCRIPTION FACTOR.
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=B;
 CC IsoId=Q9UL36-1; Sequence=displayed;
 CC Name=A;
 CC IsoId=Q9UL36-2; Sequence=VSP_006907, VSP_006908;
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS. EXPRESSION LEVELS ARE HIGHEST IN
 CC SKELETAL MUSCLE AND BRAIN, INTERMEDIATE IN HEART, PANCREAS, AND
 CC PLACENTA, AND LOWEST IN KIDNEY, LIVER, AND LUNG.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF085244; AAD55329.1; -;
 CC EMBL: AF085243; AAD55328.1; -;
 CC HSSP: P07246; IARD.
 CC GeneW: HGNC:13028; ZNF236.
 CC MIM: 604760; -;
 CC DR GO: GO:0005634; C:nucleus; NAS.
 CC DR GO: GO:0003700; F:transcription factor activity; NAS.
 CC DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC DR InterPro: IPR007087; Znf_C2H2.
 CC DR Pfam: PF00096; Zf_C2H2; 30.
 CC DR ProDom: PD000003; Znf_C2H2; 4.
 CC DR SMART: SM00355; Znf_C2H2; 30.
 CC DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 30.
 CC DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 30.
 CC DR Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 CC Nuclear protein; Repeat; Alternative splicing.
 CC KW Nuclear protein; Repeat; Alternative splicing.
 CC FT ZN_FING 37 59
 CC FT ZN_FING 66 88 C2H2-TYPE.
 CC FT ZN_FING 93 115 C2H2-TYPE.
 CC FT ZN_FING 121 143 C2H2-TYPE.
 CC FT ZN_FING 153 175 C2H2-TYPE.
 CC FT ZN_FING 197 219 C2H2-TYPE.
 CC FT ZN_FING 225 247 C2H2-TYPE.
 CC FT ZN_FING 253 276 C2H2-TYPE.
 CC FT ZN_FING 285 308 C2H2-TYPE.
 CC FT ZN_FING 482 504 C2H2-TYPE.
 CC FT ZN_FING 510 532 C2H2-TYPE.
 CC FT ZN_FING 538 560 C2H2-TYPE.
 CC FT ZN_FING 566 588 C2H2-TYPE.
 CC FT ZN_FING 657 679 C2H2-TYPE.
 CC FT ZN_FING 685 707 C2H2-TYPE.
 CC FT ZN_FING 713 735 C2H2-TYPE.
 CC FT ZN_FING 741 763 C2H2-TYPE.
 CC FT ZN_FING 967 989 C2H2-TYPE.
 CC FT ZN_FING 995 1017 C2H2-TYPE.
 CC FT ZN_FING 1023 1045 C2H2-TYPE.
 CC FT ZN_FING 1051 1073 C2H2-TYPE.
 CC FT ZN_FING 1166 1189 C2H2-TYPE.
 CC FT ZN_FING 1167 1189 C2H2-TYPE.
 CC FT ZN_FING 1195 1217 C2H2-TYPE.
 CC FT ZN_FING 1223 1245 C2H2-TYPE.
 CC FT ZN_FING 1251 1273 C2H2-TYPE.
 CC FT ZN_FING 1657 1680 C2H2-TYPE.
 CC FT ZN_FING 1686 1708 C2H2-TYPE.
 CC FT ZN_FING 1722 1744 C2H2-TYPE.

FT ZN_FING 1750 1772 C2H2-TYPE.
 FT ZN_FING 1778 1801 C2H2-TYPE.
 FT VARSPPLIC 1530 1558 ELNATTSGLPSTPTPSPSAISTONLWMS -> GSRVQHS
 FT VARSPPLIC 1530 1558 VQPCGSAVEALYENLSBKT (in isoform A).
 FT VARSPPLIC 1559 1845 Missing (in isoform A).
 FT VARSPPLIC 1559 1845 /FTID=VSP_006908.
 SQ SEQUENCE 1845 AA; 203659 MW; 2879EA91D0C6DD3D8 CRC64;
 Query Match 8.2%; Score 407; DB 1; Length 1845;
 Best local similarity 20.6%; Pred. No. 5.8e-12;
 Matches 220; Conservative 133; Mismatches 385; Indels 370; Gaps 49;
 QY 7 GNRKFKTECGAKFKYKHLKEHLRHSGEKPECPNCKRFSHSGSYSHSKKICGL 66
 DB 221 GERPFKSCGAFNKGALQTHMIKHTEKHACFCFAASQKGNLQSHVQ----- 273
 QY 67 ISVNGMRNNITGSSPNSVSSPTNSALTQIR--NKLNGKPLMSQTLGK---- 119
 DB 274 -----RVHSEVKNQPTYNCTSCSCVFKSLGSLNTHISKMHGQPOSTSTETAHLVAT 328
 QY 120 --KTEPLDNDYKVLMTATGFGSG--SPMNGGLATSPLVHPASQSPM--QHLG--VG 171
 DB 329 LQQTLPLOQTEAATASASQPSQAVSDVIQQLLESEAPVE--SQSPQPGQSLITVG 387
 QY 172 MEAPLLGFPTMNSNLSEVQKVLQIVDNTVSRQKMDCKEDISKLGYNHMKDPC---SQP 227
 DB 388 INDILQALENSGLSIPAA-----HPNDSCHAKTSAP 422
 QY 228 EERGVTSPNIPYGVLPVSHNGATSIIDYLEKNVEAKCIQSLTDSRQISNKKER 287
 DB 423 HQ--NPDVSVS-----NE-----QTPDTEOEKEDESPERK 453
 QY 288 LRTLIDLVDDKMIENHSISTPFS-----CQFKESFPGP-----IPLHQR 330
 DB 454 LDK-----KEKMIKKSPFLPGSIRENGVMHVCYAKFRKPSDLVRIHITHRK 508
 QY 331 YL-----CKNNEIKAVLQPHENIVPNKAGVFNKALLISVLESG-----LTSPI 379
 DB 509 PFKPCQCFPAFAVKSTLTGAI--IKTHGIIKAFKQCYCKMS--FSTGSLKVIHLRTGVR 564
 QY 380 PYDQHSVLKAYANMMENSDCLKIYAVGLPQEFVEMFQKRYOYVSNRSRSLR 439
 DB 565 PFAC-----PHCDKPFST--GHRKTHASHKHTELRMRQORAKYR 607
 QY 440 TSXPLAPNSPTTKOSILPRSPVFKPMDISITSPSIAE-----LHNSVSCDPL 487
 DB 608 VKGTNP-----VPDIPLOEPILITDGLIPIPKNQFPQSYFNNNFVNEADRY 657
 QY 488 R-----LTKSHFTNIKAVDKLHRSNT--PSPINLS-----STSSKNSHSSSTYP 532
 DB 658 KCFYCHRAVYKSKCHK-----QHIRSHTEKPKKSCQGRGFVSAGVLKAIHITHG 709
 QY 533 -NSFS-----SEEL-----QAEPLDLSLPQMBEPKGIAT 562
 DB 710 LKSFKCLIONGAFETTGSLRRHMGIHNDLRPYMCPYCOQTFPTSLNCKKMHKTHRELAQ 769
 QY 563 KNTTKATSLINDNSVSSSENSDEPLNTFTIKKEPSNNLNKNNPVFGMP----- 617
 DB 770 QLOQHQAASIDISTYDOOSMQASTQMOVEIESDELPTAEVVAANPEAMLDLEPHVVG 829
 QY 618 -----FSAKPL-----YTPL-----PQSAFPPA-----T 637
 DB 830 TEEAGLGOQLADQPLBADBDGFVAPDPRLGHVDCFEESQSPAQOSPEPAGLPQGFVTDT 889
 QY 638 F-----MPVQ-----TSI--PGLRPTFGLDQMSFLPHMAVYPTGATRA 676
 DB 890 YHQQPPFPVQQLQDSSTLESQALSTSFHQSLQLQAPSSDGANNVTRL-----IQ 939
 QY 677 DMQQRKRYRKQGFQDLDLDAQDVYSGLDMDTDSCLSRKKIKKTESGMYACDLCKT 736
 DB 940 ESSQEBELDLQAQSQ-----FLEDNEDQ-----SRNS-----YRCDICNKG 975

QY	737	FOKSSSLRHHKXETGKPPHOCQICKAFKPKHHLIEHSLSGCEPKVQCCKGRFSHS	796
Db	976	FKSSHLKQHRSHRTGEPKPKCKLCGRFVSGLKSHKHTHYKAFSCVCNASTFTN	10355
QY	797	GSYSQHM-----NHRYSYCKREAEREAREAREAREKHLGPTLLMNRAYLOSITPOG	849
Db	1036	GSLLRRHMAATHWSMKPKYKCFCE-----EGPRTTYHCKKH-----MKR---HQTVSA	1079
QY	850	YSDBEERSMNRDSESEKEHEKEGEGYCKLRRDGDDEEBEHEESE-NKSMOTDPTI	908
Db	1080	VSATGEEEG-----GDICWEEEBEESHSDRNARSRSRPEVI	1113
QY	909	R-DEETGD-----HSMDDSSDGKMTKSHDEEDNMD	941
Db	1114	TTTEETRAQLAKIRPQESATVSEKVLVSAAKORISE	1151

RESULT 10

ID	ZN91_HUMAN	STANDARD;	PRT; 1191 AA.
AC	Q05481;		
DT	01-JUN-1994 (Rel. 29, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	Zinc finger protein 91 (Zinc finger protein HTP10) (HPF7).		
GN	ZNF91.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=93223677; PubMed=8467795;		
RA	Bellefroid E.J., Marine J.C., Ried T., Lecocq P.J., Riviere M.,		
RA	Ameijia C.T., Poncelet D.A., Coulle P.G., de Jong P.J.,		
RA	Szpirer C., Ward D.C., Martial J.A.		
RT	"Clustered organization of homologous KRAB zinc-finger genes with		
RT	enhanced expression in human T lymphoid cells."		
RL	EMBO J. 12:1363-1374 (1993).		
RN	[2]		
RP	SEQUENCE OF 15-204 FROM N.A.		
RX	MEDLINE=91219421; PubMed=2023909;		
RA	Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,		
RA	Martial J.A.;		
RT	"The evolutionarily conserved Kruppel-associated box domain defines a		
RT	subfamily of eukaryotic multifingered proteins."		
RL	Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612 (1991).		
CC	-1- SUBCELLULAR LOCATION: Nuclear (Probable).		
CC	-1- SIMILARITY: BELONGS TO THE KRUEPEL FAMILY OF C2H2-TYPE ZINC-		
CC	FINGER PROTEINS.		
CC	-1- SIMILARITY: Contains 1 KRAB domain.		
CC	-1- CAUTION: THE SEQUENCE FROM POSITION 1159 TO THE C-TERMINAL IS		
CC	DERIVED FROM THE TRANSLATION OF AN ALU REPEAT.		

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb.ch/announce/>) or send an email to license@isb-sib.ch).

CC

CC EMBL; L11672; AAA59469.1; -;

DR EMBL; M61871; AAA56672.1; ALT_SEQ.

DR PIR; S35305; S35305.

DR HSSP; P08047; ISP2.

DR Genew; HGNC:13166; ZNF91.

DR MIM; 603971; -;

DR GO; GO:0005634; C:nucleus; NAS.

DR GO; GO:0003700; F:transcription factor activity; NAS.

DR GO; GO:0008270; F:zinc ion binding activity; NAS.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.

DR InterPro; IPR001909; KRAB.

DR	InterPro:	IPR007087;	Znf_C2H2.
DR	InterPro:	IPR007086;	Znf_C2H2_sub.
DR	pfam;	PF01352;	KRAB; 1.
DR	pfam;	Pf00906;	zf-C2H2; 34.
DR	PRINTS;	PR00048;	ZINC_FINGER.
DR	ProDom;	PD000003;	Znf_C2H2; 20.
DR	SMART;	SM00349;	KRAB; 1.
DR	SMART;	SM00355;	Znf_C2H2; 33.
DR	PROSITE;	PS50805;	KRAB; 1.
DR	PROSITE;	PS00028;	ZINC_FINGER_C2H2_1; 31.
DR	PROSITE;	PS50157;	ZINC_FINGER_C2H2_2; 35.
KW	Transcription regulation;	Zinc-finger;	DNA-binding; Metal-binding;
KW	Nuclear protein;	Repeat.	
FT	DOMAIN	13	KRAB
FT	ZN_FING	154	176 C2H2-TYPE (DEGENERATE).
FT	ZN_FING	179	200 C2H2-TYPE.
FT	ZN_FING	208	232 C2H2-TYPE.
FT	ZN_FING	238	260 C2H2-TYPE.
FT	ZN_FING	266	288 C2H2-TYPE.
FT	ZN_FING	294	316 C2H2-TYPE.
FT	ZN_FING	322	344 C2H2-TYPE.
FT	ZN_FING	350	372 C2H2-TYPE.
FT	ZN_FING	378	400 C2H2-TYPE.
FT	ZN_FING	406	428 C2H2-TYPE.
FT	ZN_FING	434	456 C2H2-TYPE.
FT	ZN_FING	462	484 C2H2-TYPE.
FT	ZN_FING	490	512 C2H2-TYPE.
FT	ZN_FING	518	540 C2H2-TYPE (DEGENERATE).
FT	ZN_FING	546	568 C2H2-TYPE.
FT	ZN_FING	574	596 C2H2-TYPE.
FT	ZN_FING	602	624 C2H2-TYPE.
FT	ZN_FING	630	652 C2H2-TYPE.
FT	ZN_FING	658	680 C2H2-TYPE.
FT	ZN_FING	686	708 C2H2-TYPE.
FT	ZN_FING	714	736 C2H2-TYPE.
FT	ZN_FING	742	764 C2H2-TYPE.
FT	ZN_FING	770	792 C2H2-TYPE.
FT	ZN_FING	798	820 C2H2-TYPE.
FT	ZN_FING	826	848 C2H2-TYPE.
FT	ZN_FING	854	876 C2H2-TYPE.
FT	ZN_FING	885	904 C2H2-TYPE (DEGENERATE).
FT	ZN_FING	904	932 C2H2-TYPE.
FT	ZN_FING	938	960 C2H2-TYPE.
FT	ZN_FING	966	988 C2H2-TYPE.
FT	ZN_FING	994	1016 C2H2-TYPE.
FT	ZN_FING	1022	1044 C2H2-TYPE.
FT	ZN_FING	1050	1072 C2H2-TYPE.
FT	ZN_FING	1078	1100 C2H2-TYPE.
FT	ZN_FING	1106	1128 C2H2-TYPE.
FT	ZN_FING	1134	1156 C2H2-TYPE.
SQ	SEQUENCE	1191 AA;	137225 MW; 581056BB1B8716D CRC64;
Query Match 7.6%; Score 379; DB 1; Length 1191;			
Best Local Similarity 21.5%; Pred. No. 7.1e-11;			
Matches 214; Conservative 115; Mismatches 365; Indels 300; Gaps 39			
OY	7	GNRKFKTECCAKAFKYKHHLKEHLRIHSGEYPCEGPNCKKSPSHSGSYSHISXKCIGL	66
Db	290	GEKYKCECGCAKASHSTLAKRKIRIHGEKPYKCECGKAPSHSALAKH-----	340
OY	67	ISVNGRMNNIKTGSSPN-----SVSSPTNSAITOLRNK-----	101
Db	341	-----KRHTGEKPYKCECGKAFSNSSTLANHKITHHEKPYKCECDTKFRLST	392
OY	102	-----LENGRPILMSBOTGLKLTETELDNDNYKVLMATTCF--SGSSPFWNGLGA--	151
Db	393	LTRKHIITHAAGEKLKYKCECGA-----ENRSSNL-TIHFHTHGEKPYKCECGKAF	443
OY	152	--TSPLGVHP---SAOSPWOHAGVMEADPLLGFPMTNSNLSEVKVLQIVDTVSROKMD	206
Db	444	NWSSSLTKHKRFHIREKPKCECGK-----GF-----IMSTLTRHKRI	483
OY	207	CKTEDISKLKGYHMKDPCSQPEEGCVTSPIPIPVGLPVVSHNGATYSIIIDVTLEKVNEMAK	266

```

Db 484 HNGE-----KPYKCEB-CGKAFROSST-----LTKHKIHTHEKPYKCE 521
Qy 267 ACLOSITDSRQISINIKKELATLIDLVTDDMIENHSITPSPCOFCSEPGPIPLH 326
Db 522 ECGKAF-----KOSLTLNKHKI-----IHSKPYKCECGKAKKOSTLT 562
Qy 327 QHE-----RYLCKMNEIKAV-----LOPHENIVPKAGVVD--NKALLSSVL--SEKG 373
Db 563 THKIIHAKGKLYKCECGKAFNHSLSLTHKIIHTGEKSKYKCECGKAFWSLTLRHKR 622
Qy 374 LNSPIPKYK-----DHMSVLKAYAMMENBELKISTAVGLPQFVKEWEOQR 425
Db 623 IHGKPYKCECGKAFSHSALAKH-----KRIHTGEK 656
Qy 426 VVO-----YSNRSPLERTS-----KPLAPNSPTTKOSLLPRSPVKMDSITPSJA 474
Db 657 PYKCECGKAFSSSTLANHKTHTBEKPYKCECDKT-----FKRLSTLTHKII 707
Qy 475 ELHNSVYSCDP-PLRLTKSSHPTNIRKAVDLHSRSNTPSPPLNLSSTSKNSHSSVTPN 533
Db 708 HAGEKLYKCECGKAFNRSSNLT-----IHKFIHTGE--KPYKCECGKAFNHSLSLTHK 760
Qy 534 S--FSSEBLOAEPLDSLPRKQNR-PRGIIATKNTKATSIINDHNSVSSSSSENSDEPLN 590
Db 761 KRHTREK-----PFKCECGKAFWSLTLRHKRIH-----TGEKPYK 799
Qy 591 LRTIKKEFSNSNLDKNSNPNVFGMNPFAKPLTYPLPQSPAPPAFMPVOTSIPLGR 650
Db 800 CEECGKAFSSSTLT-TKAKTIHTGEKPYKCEC-----GKAKHSSALAKHKIHTHAGEK 852
Qy 651 PY-----PGLDQMSFLPHNAVYTPGAAFPAD-----MOQRKRYRKQGFQ 692
Db 853 LYKCECGKAFNQSSTLTHKIIHTKEKPKSECDKAFWSLTLRHKIHTREKPYK 912
Qy 693 DLDGQODWISGL-----DMTDSLSLKRKIKTSSGYACDLC 734
Db 913 BECGKAFSOPSHLTTHKRMHTGKPYKCECGKAFSSSTLTHKIIHTGEKPYKCEBG 972
Qy 735 KTFQKSSSLRHYEHTGKRPKOCQCKKAFKHKHIIHSRLHSGEKPYQCDKCGKRF 794
Db 973 KAFKSSLTLEHKTHTGEKPYKCECGKAFSSSTLTTHKRMHTGKPYKCECGKAFN 1032
Qy 795 HSGSVSOHMHRYSYCKREAEEREAERAREKGLAPTELLNRAVLOSITPOGYSDE 854
Db 1033 RSEKLTTHKIIHTGEKPYKCEB-----CGKAFISSTLNGHKRIH 1072
Qy 855 ERESMPRDGSEK-----EHK--EGEGY 877
Db 1073 TREKPYKCECGKAFSSSTLTHKRLHTGEKPY 1106

RESULT 11
EV11_HUMAN STANDARD; PRT; 1051 AA.
ID EV11_HUMAN
AC 003112; Q16122; Q99917;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ecotopic virus integration 1 site protein.
GN EV11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90326419; PubMed=2115646;
RA Morihita K., Parganas E., Douglass E.C., Ihle J.N.;
RT "Unique expression of the human Evi-1 gene in an endometrial
carcinoma cell line: sequence of cDNAs and structure of alternatively
spliced transcripts."
RL Oncogene 5:963-971(1990).

```

```

RN [2]
RP SEQUENCE OF 970-1012 FROM N.A.
RX MEDLINE=96292241; PubMed=8700545;
RA Ogawa S., Kurokawa M., Tanaka T., Mitani K., Inazawa J.,
RA Hangishi A., Tanaka K., Matsuo Y., Minowada J., Tsubota T.,
RA Yasaki Y., Hirai H.;
RT "Structurally altered Evi-1 protein generated in the 3q21q26
syndrome."
RL Oncogene 13:183-191(1996).
RN [3]
RP VARIANT AML1-EVI1 FUSION IN CHRONIC MYELOCYTIC LEUKEMIA.
RX MEDLINE=94147997; PubMed=8313895;
RA Mitani K., Ogawa S., Tanaka T., Miyoshi H., Kurokawa M., Mano H.,
RA Yasaki Y., Ohki M., Hirai H.;
RT "Generation of the AML1-EVI-1 fusion gene in the t(3;21)(q26;q22)
RT causes blast crisis in chronic myelocytic leukemia."
RL EMBL J. 13:504-510(1994).
CC -1- SUBUNIT: MAY INTERACT WITH CTBP1 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=Long;
CC IsoId=Q03112-1; Sequence=Displayed;
CC Name=2; Synonyms=Short;
CC IsoId=Q03112-2; Sequence=Not described;
CC -1- DISEASE: INVOLVED IN CHRONIC MYELOGENOUS LEUKEMIA (CML) BY A
CC CHROMOSOMAL TRANSLOCATION T(3;21)(Q26;Q22) THAT INVOLVES EVI1 AND
CC PROTEIN AML1.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54989; CAA8735.1; -
DR EMBL; S82592; AAB37456.1; -
DR EMBL; S69002; AAB32907.1; ALT_SEQ.
DR PIR; A60191; A60191.
DR PIR; S41705; S41705.
DR HSSP; P08047; ISP2.
DR GeneW; HGNC:3498; EVI1.
DR MIM; 165215; -
DR GO; GO:0005634; C:nucleus; NAS.
DR InterPro; IPR007087; Znf.C2H2.
DR Pfam; PF00006; zf.C2H2; 10.
DR ProDom; PD000003; Znf.C2H2; 4.
DR SMART; SM00355; Znf.C2H2; 10.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 10.
KW Zinc-finger; Metal-binding; Developmental protein;
KW Proto-oncogene; Chromosomal translocation; Nuclear protein; Repeat;
KW Alternative splicing.
FT ZN_FING 21
FT ZN_FING 75
FT ZN_FING 103
FT ZN_FING 131
FT ZN_FING 160
FT ZN_FING 182
FT ZN_FING 217
FT ZN_FING 239
FT ZN_FING 733
FT ZN_FING 761
FT ZN_FING 764
FT ZN_FING 790
FT ZN_FING 812
FT ZN_FING 832
FT ZN_FING 844
FT ZN_FING 866
FT ZN_FING 886
FT ZN_FING 937
FT ZN_FING 957
FT ZN_FING 972
FT ZN_FING 984
FT ZN_FING 988
FT ZN_FING 993
FT ZN_FING 998
FT ZN_FING 1003
FT ZN_FING 1008
FT ZN_FING 1012
FT ZN_FING 1017
FT ZN_FING 1022
FT ZN_FING 1027
FT ZN_FING 1032
FT ZN_FING 1037
FT ZN_FING 1042
FT ZN_FING 1047
FT ZN_FING 1051
FT ZN_FING 1056
FT ZN_FING 1061
FT ZN_FING 1066
FT ZN_FING 1071
FT ZN_FING 1076
FT ZN_FING 1081
FT ZN_FING 1086
FT ZN_FING 1091
FT ZN_FING 1096
FT ZN_FING 1101
FT ZN_FING 1106
FT ZN_FING 1111
FT ZN_FING 1116
FT ZN_FING 1121
FT ZN_FING 1126
FT ZN_FING 1131
FT ZN_FING 1136
FT ZN_FING 1141
FT ZN_FING 1146
FT ZN_FING 1151
FT ZN_FING 1156
FT ZN_FING 1161
FT ZN_FING 1166
FT ZN_FING 1171
FT ZN_FING 1176
FT ZN_FING 1181
FT ZN_FING 1186
FT ZN_FING 1191
FT ZN_FING 1196
FT ZN_FING 1201
FT ZN_FING 1206
FT ZN_FING 1211
FT ZN_FING 1216
FT ZN_FING 1221
FT ZN_FING 1226
FT ZN_FING 1231
FT ZN_FING 1236
FT ZN_FING 1241
FT ZN_FING 1246
FT ZN_FING 1251
FT ZN_FING 1256
FT ZN_FING 1261
FT ZN_FING 1266
FT ZN_FING 1271
FT ZN_FING 1276
FT ZN_FING 1281
FT ZN_FING 1286
FT ZN_FING 1291
FT ZN_FING 1296
FT ZN_FING 1301
FT ZN_FING 1306
FT ZN_FING 1311
FT ZN_FING 1316
FT ZN_FING 1321
FT ZN_FING 1326
FT ZN_FING 1331
FT ZN_FING 1336
FT ZN_FING 1341
FT ZN_FING 1346
FT ZN_FING 1351
FT ZN_FING 1356
FT ZN_FING 1361
FT ZN_FING 1366
FT ZN_FING 1371
FT ZN_FING 1376
FT ZN_FING 1381
FT ZN_FING 1386
FT ZN_FING 1391
FT ZN_FING 1396
FT ZN_FING 1401
FT ZN_FING 1406
FT ZN_FING 1411
FT ZN_FING 1416
FT ZN_FING 1421
FT ZN_FING 1426
FT ZN_FING 1431
FT ZN_FING 1436
FT ZN_FING 1441
FT ZN_FING 1446
FT ZN_FING 1451
FT ZN_FING 1456
FT ZN_FING 1461
FT ZN_FING 1466
FT ZN_FING 1471
FT ZN_FING 1476
FT ZN_FING 1481
FT ZN_FING 1486
FT ZN_FING 1491
FT ZN_FING 1496
FT ZN_FING 1501
FT ZN_FING 1506
FT ZN_FING 1511
FT ZN_FING 1516
FT ZN_FING 1521
FT ZN_FING 1526
FT ZN_FING 1531
FT ZN_FING 1536
FT ZN_FING 1541
FT ZN_FING 1546
FT ZN_FING 1551
FT ZN_FING 1556
FT ZN_FING 1561
FT ZN_FING 1566
FT ZN_FING 1571
FT ZN_FING 1576
FT ZN_FING 1581
FT ZN_FING 1586
FT ZN_FING 1591
FT ZN_FING 1596
FT ZN_FING 1601
FT ZN_FING 1606
FT ZN_FING 1611
FT ZN_FING 1616
FT ZN_FING 1621
FT ZN_FING 1626
FT ZN_FING 1631
FT ZN_FING 1636
FT ZN_FING 1641
FT ZN_FING 1646
FT ZN_FING 1651
FT ZN_FING 1656
FT ZN_FING 1661
FT ZN_FING 1666
FT ZN_FING 1671
FT ZN_FING 1676
FT ZN_FING 1681
FT ZN_FING 1686
FT ZN_FING 1691
FT ZN_FING 1696
FT ZN_FING 1701
FT ZN_FING 1706
FT ZN_FING 1711
FT ZN_FING 1716
FT ZN_FING 1721
FT ZN_FING 1726
FT ZN_FING 1731
FT ZN_FING 1736
FT ZN_FING 1741
FT ZN_FING 1746
FT ZN_FING 1751
FT ZN_FING 1756
FT ZN_FING 1761
FT ZN_FING 1766
FT ZN_FING 1771
FT ZN_FING 1776
FT ZN_FING 1781
FT ZN_FING 1786
FT ZN_FING 1791
FT ZN_FING 1796
FT ZN_FING 1801
FT ZN_FING 1806
FT ZN_FING 1811
FT ZN_FING 1816
FT ZN_FING 1821
FT ZN_FING 1826
FT ZN_FING 1831
FT ZN_FING 1836
FT ZN_FING 1841
FT ZN_FING 1846
FT ZN_FING 1851
FT ZN_FING 1856
FT ZN_FING 1861
FT ZN_FING 1866
FT ZN_FING 1871
FT ZN_FING 1876
FT ZN_FING 1881
FT ZN_FING 1886
FT ZN_FING 1891
FT ZN_FING 1896
FT ZN_FING 1901
FT ZN_FING 1906
FT ZN_FING 1911
FT ZN_FING 1916
FT ZN_FING 1921
FT ZN_FING 1926
FT ZN_FING 1931
FT ZN_FING 1936
FT ZN_FING 1941
FT ZN_FING 1946
FT ZN_FING 1951
FT ZN_FING 1956
FT ZN_FING 1961
FT ZN_FING 1966
FT ZN_FING 1971
FT ZN_FING 1976
FT ZN_FING 1981
FT ZN_FING 1986
FT ZN_FING 1991
FT ZN_FING 1996
FT ZN_FING 2001
FT ZN_FING 2006
FT ZN_FING 2011
FT ZN_FING 2016
FT ZN_FING 2021
FT ZN_FING 2026
FT ZN_FING 2031
FT ZN_FING 2036
FT ZN_FING 2041
FT ZN_FING 2046
FT ZN_FING 2051
FT ZN_FING 2056
FT ZN_FING 2061
FT ZN_FING 2066
FT ZN_FING 2071
FT ZN_FING 2076
FT ZN_FING 2081
FT ZN_FING 2086
FT ZN_FING 2091
FT ZN_FING 2096
FT ZN_FING 2101
FT ZN_FING 2106
FT ZN_FING 2111
FT ZN_FING 2116
FT ZN_FING 2121
FT ZN_FING 2126
FT ZN_FING 2131
FT ZN_FING 2136
FT ZN_FING 2141
FT ZN_FING 2146
FT ZN_FING 2151
FT ZN_FING 2156
FT ZN_FING 2161
FT ZN_FING 2166
FT ZN_FING 2171
FT ZN_FING 2176
FT ZN_FING 2181
FT ZN_FING 2186
FT ZN_FING 2191
FT ZN_FING 2196
FT ZN_FING 2201
FT ZN_FING 2206
FT ZN_FING 2211
FT ZN_FING 2216
FT ZN_FING 2221
FT ZN_FING 2226
FT ZN_FING 2231
FT ZN_FING 2236
FT ZN_FING 2241
FT ZN_FING 2246
FT ZN_FING 2251
FT ZN_FING 2256
FT ZN_FING 2261
FT ZN_FING 2266
FT ZN_FING 2271
FT ZN_FING 2276
FT ZN_FING 2281
FT ZN_FING 2286
FT ZN_FING 2291
FT ZN_FING 2296
FT ZN_FING 2301
FT ZN_FING 2306
FT ZN_FING 2311
FT ZN_FING 2316
FT ZN_FING 2321
FT ZN_FING 2326
FT ZN_FING 2331
FT ZN_FING 2336
FT ZN_FING 2341
FT ZN_FING 2346
FT ZN_FING 2351
FT ZN_FING 2356
FT ZN_FING 2361
FT ZN_FING 2366
FT ZN_FING 2371
FT ZN_FING 2376
FT ZN_FING 2381
FT ZN_FING 2386
FT ZN_FING 2391
FT ZN_FING 2396
FT ZN_FING 2401
FT ZN_FING 2406
FT ZN_FING 2411
FT ZN_FING 2416
FT ZN_FING 2421
FT ZN_FING 2426
FT ZN_FING 2431
FT ZN_FING 2436
FT ZN_FING 2441
FT ZN_FING 2446
FT ZN_FING 2451
FT ZN_FING 2456
FT ZN_FING 2461
FT ZN_FING 2466
FT ZN_FING 2471
FT ZN_FING 2476
FT ZN_FING 2481
FT ZN_FING 2486
FT ZN_FING 2491
FT ZN_FING 2496
FT ZN_FING 2501
FT ZN_FING 2506
FT ZN_FING 2511
FT ZN_FING 2516
FT ZN_FING 2521
FT ZN_FING 2526
FT ZN_FING 2531
FT ZN_FING 2536
FT ZN_FING 2541
FT ZN_FING 2546
FT ZN_FING 2551
FT ZN_FING 2556
FT ZN_FING 2561
FT ZN_FING 2566
FT ZN_FING 2571
FT ZN_FING 2576
FT ZN_FING 2581
FT ZN_FING 2586
FT ZN_FING 2591
FT ZN_FING 2596
FT ZN_FING 2601
FT ZN_FING 2606
FT ZN_FING 2611
FT ZN_FING 2616
FT ZN_FING 2621
FT ZN_FING 2626
FT ZN_FING 2631
FT ZN_FING 2636
FT ZN_FING 2641
FT ZN_FING 2646
FT ZN_FING 2651
FT ZN_FING 2656
FT ZN_FING 2661
FT ZN_FING 2666
FT ZN_FING 2671
FT ZN_FING 2676
FT ZN_FING 2681
FT ZN_FING 2686
FT ZN_FING 2691
FT ZN_FING 2696
FT ZN_FING 2701
FT ZN_FING 2706
FT ZN_FING 2711
FT ZN_FING 2716
FT ZN_FING 2721
FT ZN_FING 2726
FT ZN_FING 2731
FT ZN_FING 2736
FT ZN_FING 2741
FT ZN_FING 2746
FT ZN_FING 2751
FT ZN_FING 2756
FT ZN_FING 2761
FT ZN_FING 2766
FT ZN_FING 2771
FT ZN_FING 2776
FT ZN_FING 2781
FT ZN_FING 2786
FT ZN_FING 2791
FT ZN_FING 2796
FT ZN_FING 2801
FT ZN_FING 2806
FT ZN_FING 2811
FT ZN_FING 2816
FT ZN_FING 2821
FT ZN_FING 2826
FT ZN_FING 2831
FT ZN_FING 2836
FT ZN_FING 2841
FT ZN_FING 2846
FT ZN_FING 2851
FT ZN_FING 2856
FT ZN_FING 2861
FT ZN_FING 2866
FT ZN_FING 2871
FT ZN_FING 2876
FT ZN_FING 2881
FT ZN_FING 2886
FT ZN_FING 2891
FT ZN_FING 2896
FT ZN_FING 2901
FT ZN_FING 2906
FT ZN_FING 2911
FT ZN_FING 2916
FT ZN_FING 2921
FT ZN_FING 2926
FT ZN_FING 2931
FT ZN_FING 2936
FT ZN_FING 2941
FT ZN_FING 2946
FT ZN_FING 2951
FT ZN_FING 2956
FT ZN_FING 2961
FT ZN_FING 2966
FT ZN_FING 2971
FT ZN_FING 2976
FT ZN_FING 2981
FT ZN_FING 2986
FT ZN_FING 2991
FT ZN_FING 2996
FT ZN_FING 3001
FT ZN_FING 3006
FT ZN_FING 3011
FT ZN_FING 3016
FT ZN_FING 3021
FT ZN_FING 3026
FT ZN_FING 3031
FT ZN_FING 3036
FT ZN_FING 3041
FT ZN_FING 3046
FT ZN_FING 3051
FT ZN_FING 3056
FT ZN_FING 3061
FT ZN_FING 3066
FT ZN_FING 3071
FT ZN_FING 3076
FT ZN_FING 3081
FT ZN_FING 3086
FT ZN_FING 3091
FT ZN_FING 3096
FT ZN_FING 3101
FT ZN_FING 3106
FT ZN_FING 3111
FT ZN_FING 3116
FT ZN_FING 3121
FT ZN_FING 3126
FT ZN_FING 3131
FT ZN_FING 3136
FT ZN_FING 3141
FT ZN_FING 3146
FT ZN_FING 3151
FT ZN_FING 3156
FT ZN_FING 3161
FT ZN_FING 3166
FT ZN_FING 3171
FT ZN_FING 3176
FT ZN_FING 3181
FT ZN_FING 3186
FT ZN_FING 3191
FT ZN_FING 3196
FT ZN_FING 3201
FT ZN_FING 3206
FT ZN_FING 3211
FT ZN_FING 3216
FT ZN_FING 3221
FT ZN_FING 3226
FT ZN_FING 3231
FT ZN_FING 3236
FT ZN_FING 3241
FT ZN_FING 3246
FT ZN_FING 3251
FT ZN_FING 3256
FT ZN_FING 3261
FT ZN_FING 3266
FT ZN_FING 3271
FT ZN_FING 3276
FT ZN_FING 3281
FT ZN_FING 3286
FT ZN_FING 3291
FT ZN_FING 3296
FT ZN_FING 3301
FT ZN_FING 3306
FT ZN_FING 3311
FT ZN_FING 3316
FT ZN_FING 3321
FT ZN_FING 3326
FT ZN_FING 3331
FT ZN_FING 3336
FT ZN_FING 3341
FT ZN_FING 3346
FT ZN_FING 3351
FT ZN_FING 3356
FT ZN_FING 3361
FT ZN_FING 3366
FT ZN_FING 3371
FT ZN_FING 3376
FT ZN_FING 3381
FT ZN_FING 3386
FT ZN_FING 3391
FT ZN_FING 3396
FT ZN_FING 3401
FT ZN_FING 3406
FT ZN_FING 3411
FT ZN_FING 3416
FT ZN_FING 3421
FT ZN_FING 3426
FT ZN_FING 3431
FT ZN_FING 3436
FT ZN_FING 3441
FT ZN_FING 3446
FT ZN_FING 3451
FT ZN_FING 3456
FT ZN_FING 3461
FT ZN_FING 3466
FT ZN_FING 3471
FT ZN_FING 3476
FT ZN_FING 3481
FT ZN_FING 3486
FT ZN_FING 3491
FT ZN_FING 3496
FT ZN_FING 3501
FT ZN_FING 3506
FT ZN_FING 3511
FT ZN_FING 3516
FT ZN_FING 3521
FT ZN_FING 3526
FT ZN_FING 3531
FT ZN_FING 3536
FT ZN_FING 3541
FT ZN_FING 3546
FT ZN_FING 3551
FT ZN_FING 3556
FT ZN_FING 3561
FT ZN_FING 3566
FT ZN_FING 3571
FT ZN_FING 3576
FT ZN_FING 3581
FT ZN_FING 3586
FT ZN_FING 3591
FT ZN_FING 3596
FT ZN_FING 3601
FT ZN_FING 3606
FT ZN_FING 3611
FT ZN_FING 3616
FT ZN_FING 3621
FT ZN_FING 3626
FT ZN_FING 3631
FT ZN_FING 3636
FT ZN_FING 3641
FT ZN_FING 3646
FT ZN_FING 3651
FT ZN_FING 3656
FT ZN_FING 3661
FT ZN_FING 3666
FT ZN_FING 3671
FT ZN_FING 3676
FT ZN_FING 3681
FT ZN_FING 3686
FT ZN_FING 3691
FT ZN_FING 3696
FT ZN_FING 3701
FT ZN_FING 3706
FT ZN_FING 3711
FT ZN_FING 3716
FT ZN_FING 3721
FT ZN_FING 3726
FT ZN_FING 3731
FT ZN_FING 3736
FT ZN_FING 3741
FT ZN_FING 3746
FT ZN_FING 3751
FT ZN_FING 3756
FT ZN_FING 3761
FT ZN_FING 3766
FT ZN_FING 3771
FT ZN_FING 3776
FT ZN_FING 3781
FT ZN_FING 3786
FT ZN_FING 3791
FT ZN_FING 3796
FT ZN_FING 3801
FT ZN_FING 3806
FT ZN_FING 3811
FT ZN_FING 3816
FT ZN_FING 3821
FT ZN_FING 3826
FT ZN_FING 3831
FT ZN_FING 3836
FT ZN_FING 3841
FT ZN_FING 3846
FT ZN_FING 3851
FT ZN_FING 3856
FT ZN_FING 3861
FT ZN_FING 3866
FT ZN_FING 3871
FT ZN_FING 3876
FT ZN_FING 3881
FT ZN_FING 3886
FT ZN_FING 3891
FT ZN_FING 3896
FT ZN_FING 3901
FT ZN_FING 3906
FT ZN_FING 3911
FT ZN_FING 3916
FT ZN_FING 3921
FT ZN_FING 3926
FT ZN_FING 3931
FT ZN_FING 3936
FT ZN_FING 3941
FT ZN_FING 3946
FT ZN_FING 3951
FT ZN_FING 3956
FT ZN_FING 3961
FT ZN_FING 3966
FT ZN_FING 3971
FT ZN_FING 3976
FT ZN_FING 3981
FT ZN_FING 3986
FT ZN_FING 3991
FT ZN_FING 3996
FT ZN_FING 4001
FT ZN_FING 4006
FT ZN_FING 4011
FT ZN_FING 4016
FT ZN_FING 4021
FT ZN_FING 4026
FT ZN_FING 4031
FT ZN_FING 4036
FT ZN_FING 4041
FT ZN_FING 4046
FT ZN_FING 4051
FT ZN_FING 4056
FT ZN_FING 4061
FT ZN_FING 4066
FT ZN_FING 4071
FT ZN_FING 4076
FT ZN_FING 4081
FT ZN_FING 4086
FT ZN_FING 4091
FT ZN_FING 4096
FT ZN_FING 4101
FT ZN_FING 4106
FT ZN_FING 4111
FT ZN_FING 4116
FT ZN_FING 4121
FT ZN_FING 4126
FT ZN_FING 4131
FT ZN_FING 4136
FT ZN_FING 4141
FT ZN_FING 4146
FT ZN_FING 4151
FT ZN_FING 4156
FT ZN_FING 4161
FT ZN_FING 4166
FT ZN_FING 4171
FT ZN_FING 4176
FT ZN_FING 4181
FT ZN_FING 4186
FT ZN_FING 4191
FT ZN_FING 4196
FT ZN_FING 4201
FT ZN_FING 4206
FT ZN_FING 4211
FT ZN_FING 4216
FT ZN_FING 4221
FT ZN_FING 4226
FT ZN_FING 4231
FT ZN_FING 4236
FT ZN_FING 4241
FT ZN_FING 4246
FT ZN_FING 4251
FT ZN_FING 4256
FT ZN_FING 4261
FT ZN_FING 4266
FT ZN_FING 4271
FT ZN_FING 4276
FT ZN_FING 4281
FT ZN_FING 4286
FT ZN_FING 4291
FT ZN_FING 4296
FT ZN_FING 4301
FT ZN_FING 4306
FT ZN_FING 4311
FT ZN_FING 4316
FT ZN_FING 4321
FT ZN_FING 4326
FT ZN_FING 4331
FT ZN_FING 4336
FT ZN_FING 4341
FT ZN_FING 4346
FT ZN_FING 4351
FT ZN_FING 4356
FT ZN_FING 4361
FT ZN_FING 4366
FT ZN_FING 4371
FT ZN_FING 4376
FT ZN_FING 4381
FT ZN_FING 4386
FT ZN_FING 4391
FT ZN_FING 4396
FT ZN_FING 4401
FT ZN_FING 4406
FT ZN_FING 4411
FT ZN_FING 4416
FT ZN_FING 4421
FT ZN_FING 4426
FT ZN_FING 4431
FT ZN_FING 4436
FT ZN_FING 4441
FT ZN_FING 4446
FT ZN_FING 4451
FT ZN_FING 4456
FT ZN_FING 4461
FT ZN_FING 4466
FT ZN_FING 4471
FT ZN_FING 4476
FT ZN_FING 4481
FT ZN_FING 4486
FT ZN_FING 4491
FT ZN_FING 4496
FT ZN_FING 4501
FT ZN_FING 4506
FT ZN_FING 4511
FT ZN_FING 4516
FT ZN_FING 4521
FT ZN_FING 4526
FT ZN_FING 4531
FT ZN_FING 4536
FT ZN_FING 4541
FT ZN_FING 4546
FT ZN_FING 4551
FT ZN_FING 4556
FT ZN_FING 4561
FT ZN_FING 4566
FT ZN_FING 4571
FT ZN_FING 4576
FT ZN_FING 4581
FT ZN_FING 4586
FT ZN_FING 4591
FT ZN_FING 4596
FT ZN_FING 4601
FT ZN_FING 4606
FT ZN_FING 4611
FT ZN_FING 4616
FT ZN_FING 4621
FT ZN_FING 4626
FT ZN_FING 4631
FT ZN_FING 4636
FT ZN_FING 4641
FT ZN_FING 4646
FT ZN_FING 4651
FT ZN_FING 4656
FT ZN_FING 4661
FT ZN_FING 4666
FT ZN_FING 4671
FT ZN_FING 4676
FT ZN_FING 4681
FT ZN_FING 4686
FT ZN_FING 4691
FT ZN_FING 4696
FT ZN_FING 4701
FT ZN_FING 4706
FT ZN_FING 4711
FT ZN_FING 4716
FT ZN_FING 4721
FT ZN_FING 4726
FT ZN_FING 4731
FT ZN_FING 4736
FT ZN_FING 4741
FT ZN_FING 4746
FT ZN_FING 4751
FT ZN_FING 4756
FT ZN_FING 4761
FT ZN_FING 4766
FT ZN_FING 4771
FT ZN_FING 4776
FT ZN_FING 4781
FT ZN_FING 4786
FT ZN_FING 4791
FT ZN_FING 4796
FT ZN_FING 4801
FT ZN_FING 4806
FT ZN_FING 4811
FT ZN_FING 4816
FT ZN_FING 4821
FT ZN_FING 4826
FT ZN_FING 4831
FT ZN_FING 4836
FT ZN_FING 4841
FT ZN_FING 4846
FT ZN_FING 4851
FT ZN_FING 4856
FT ZN_FING 4861
FT ZN_FING 4866
FT ZN_FING 4871
FT ZN_FING 4876
FT ZN_FING 4881
FT ZN_FING 4886
FT ZN_FING 4891
FT ZN_FING 4896
FT ZN_FING 4901
FT ZN_FING 4906
FT ZN_FING 4911
FT ZN_FING 4916
FT ZN_FING 4921
FT ZN_FING 4926
FT ZN_FING 4931
FT ZN_FING 4936
FT ZN_FING 4941
FT ZN_FING 4946
FT ZN_FING 4951
FT ZN_FING 4956
FT ZN_FING 4961
FT ZN_FING 4966
FT ZN_FING 4971
FT ZN_FING 4976
FT ZN_FING 4981
FT ZN_FING 4986
FT ZN_FING 4991
FT ZN_FING 4996
FT ZN_FING 5001
FT ZN_FING 5006
FT ZN_FING 5011
FT ZN_FING 5016
FT ZN_FING 5021
FT ZN_FING 5026
FT ZN_FING 5031
FT ZN_FING 5036
FT ZN_FING 5041
FT ZN_FING 5046
FT ZN_FING 5051
FT ZN_FING 5056
FT ZN_FING 5061
FT ZN_FING 5066
FT ZN_FING 5071
FT ZN_FING 5076
FT ZN_FING 5081
FT ZN_FING 5086
FT ZN_FING 5091
FT ZN_FING 5096
FT ZN_FING 5101
FT ZN_FING 5106
FT ZN_FING 5111
FT ZN_FING 5116
FT ZN_FING 5121
FT ZN_FING 5126
FT ZN_FING 5131
FT ZN_FING 5136
FT ZN_FING 5141
FT ZN_FING 5146
FT ZN_FING 5151
FT ZN_FING 5156
FT ZN_FING 5161
FT ZN_FING 5166
FT ZN_FING 5171
FT ZN_FING 5176
FT ZN_FING 5181
FT ZN_FING 5186
FT ZN_FING 5191
FT ZN_FING 5196
FT ZN_FING 5201
FT ZN_FING 5206
FT ZN_FING 5211
FT ZN_FING 5216
FT ZN_FING 5221
FT ZN_FING 5226
FT ZN_FING 5231
FT ZN_FING 5236
FT ZN_FING 5241
FT ZN_FING 5246
FT ZN_FING 5251
FT ZN_FING 5256
FT ZN_FING 5261
FT ZN_FING 5266
FT ZN_FING 5271
FT ZN_FING 5276
FT ZN_FING 5281
FT ZN_FING 5286
FT ZN_FING 5291
FT ZN_FING 5296
FT ZN_FING 5301
FT ZN_FING 5306
FT ZN_FING 5311
FT ZN_FING 5316
FT ZN_FING 5321
FT ZN_FING 5326
FT ZN_FING 5331
FT ZN_FING 5336
FT ZN_FING 5341
FT ZN_FING 5346
FT ZN_FING 5351
FT ZN_FING 5356
FT ZN_FING 5361
FT ZN_FING 5366
FT ZN_FING 5371
FT ZN_FING 5376
FT ZN_FING 5381
FT ZN_FING 5386
FT ZN_FING 5391
FT ZN
```

FT CONFLICT 861 861 P -> T (IN REF. 3).
 FT CONFLICT 906 906 Y -> N (IN REF. 3).
 FT CONFLICT 1008 1012 AYAMM -> VOIEP (IN REF. 2).
 SQ SEQUENCE 1051 AA; 118335 MW; 3270955E25D99D51 CRC64;

Query Match 7.1%; Score 355.5; DB 1; Length 1051;
 Best Local Similarity 19.4%; Pred. No. 7.9e-10;

Matches 199; Conservative 140; Mismatches 395; Indels 293; Gaps 36;

9 RRFKTEGCAKFKYGHLEHRLHSGEKYECPCNCKRFRFSGSGVSHISKCCIGLIS 68
 101 REYKDCQPKAFNWSNLIRHOMSHDSKHYECENCAKFTDPSNLQRHNSQHV----- 155
 69 VNGRMNNIKTGGSSPNVSSSSPTNSAITQLRNKLKNGKPL-----SMSEQTGLIKIT 121
 156 -----GARAHACECGKTFATSSGLQKHIIHSVXPFICEVCHKSYTQSNLCRHR 208
 122 EPLD-----FRDYKVLMAITGF--SGSPFMNGCIGATSPUGVHPSAOSPMQH 167
 209 MHADRTQIKCKDCQCMFSTTSLNKHRRFCGKHHFAAGF----- 250
 168 LGVGEAPLPGPTNNSLSEVQYVLQIVDNTVSRQKDCCTEDISKLGYNMDCPSQP 227
 251 FGQGISLP--GTPAMDK-----TSMVNNSHANPGIADYFG-----ANR 286
 228 EEOGVTSNPIN-----PVGH-----PVVSHNGATKSIIDYTLKVNKAACLOSIL 273
 287 HPAGLTPTAPGFSVPGLPFGSLYHNPPLPASPVMGL--STEQNNSSQS----- 338
 274 TDSRQISNIKEKRLTLIDVTDNMIENHISITPSCQCFKESPPCGPIPLHQRHYLC 333
 339 -----PLMTHPQLP 348
 334 KMMNEIKAVLOPHENIVPNKAGVFNOKALLSVLSEKGLTPIPNYKDHMSVLKAYYA 393
 349 ATQDLTKA--LSKHPSVGDNK-----EVLQPERNSEERRPEK 384
 394 NMMEFNSDELKISTAVGLPOEFVKEWEPQKRYOVSNSRSPSLERTSKPLAPNSPTTK 453
 385 ISDQESSDLDDVSTPSCSDELT-----SGSDLESIDESDEKKEKNGKPK 432
 454 DSLPRSPVKPMD-----STSPSIAE--LHNSVTSQCPPLRLTKSSHFTNIKAVD 502
 433 DKVSPLOMLASINNKKESYNNHIFSPSLEEQTAVGAVNDISIKALASIAEKYFGSTGLVG 492
 503 KLDHRSMTPLNLSSTSSKNSHSSSYTPNFSSEELQAPBLD-----SLPQOM 553
 493 LQDKKVGALPYR--SMFPLPFPFASQSMYP--FPDRDLKSLPLKMEPOSQGEVKLQKGS 549
 554 REPKGIIATKNTKATKATINLDHNSVSSSENDEPLNLTPIKKEFSNSNLDNKNPNPVF 613
 550 SESPPDLTTKRDEKPLTPVPSKPVYATPSQDQLDLSMGRSRAASGKLTERRKNHYF 609
 614 G-----MNPFSAPLYTLPPLPSAPPATPMPV-----QTSIP-----GLRPY 652
 610 GSKKGSNVESSRASPAGSL-----QHARPTPEFMDPIYREKRLTDLPEALKEXYLPBS 663
 653 PGLDMSTLPHMAIYTPGAATFADMQRKRYQKQGFQGDLLDAQDYMSGLDMTPD-- 710
 664 PG-----FLFHFQFLPQORTWMSAIENNA--EKLESFSA--LKPASSELQSVPMFNR 715
 711 -SDSCLSRRKIKTSSGMYACDLCKTFQKSSSLRHKTEHGKRPKPHOCQKKAFFKHG 769
 716 APPNALPENLARKGE--RYTCRYCGKIFPRSANLIRHLRTHGEOYRCKYCDRFSFSS 774
 770 HLIEHSR--LHSGEKYQCDKCGKRFSGSHSGSYQHMMHRYSYCKRAEEREAERAREK 828
 775 NLQRHVRNHNKKEKPKCHLCYRCFGQQTNLDRHL-----KXHG 815
 829 HLGPELLMNRAYLOSITPGYSDBERSMPPDSEKEHEKE--GREGYKGLRRDG 885
 816 NMSTG-----ATSSPHSELSTGAILLDKEDAYFTIRNFTGNSHSGSPRNV 864

QY 866 DE-----EEEEEESENKSM---DTDEPTIRDEEETDHSMDSSDEGKNETYSOH 934
 DB 865 EERNNGSHFEKEKALVPSQNSDLDDDEVEDEVLDEE---DEBYDITGKTGKEPVTNTL 921
 QY 935 EDNMED 941
 DB 922 HEGNPD 928

RESULT 12

XFIN XENLA STANDARD; PRT; 1350 AA.
 ID P08045;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein Xfin.
 GN XFIN.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodidae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=88082679; PubMed=2826129;
 RA Ruiz I Altaba A., Perry-O'Keefe H., Melton D.A.;
 RT "Xfin: an embryonic gene encoding a multifingered protein in
 RT Xenopus.";
 RL EMBO J. 6:3065-3070(1987).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=94021366; PubMed=7692399;
 RA Andreatzoli M., de Lucchini S., Costa M., Barsacchi G.;
 RT "RNA binding properties and evolutionary conservation of the Xenopus
 RT multifinger protein Xfin.";
 RL Nucleic Acids Res. 21:4218-4225(1993).
 RN [3]
 RP STRUCTURE BY NMR OF FINGER 31.
 RX MEDLINE=89346749; PubMed=2503871;
 RA Lee M.S., Gipeert G.P., Soman K.V., Case D.A., Wright P.B.;
 RT "Three-dimensional solution structure of a single zinc finger DNA-
 RT binding domain.";
 RL Science 245:635-637(1998).
 RN [4]
 RP STRUCTURE BY NMR OF A FINGER.
 RX MEDLINE=89378224; PubMed=2506074;
 RA Lee M.S., Cavanagh J., Wright P.B.;
 RT "Complete assignment of the 1H NMR spectrum of a synthetic zinc
 RT finger from Xfin. Sequential resonance assignments and secondary
 RT structure.";
 RL FEBS Lett. 254:159-164(1989).
 CC -1- FUNCTION: BINDS RNA. COULD FUNCTION IN POST-TRANSLATIONAL
 CC REGULATION PROCESSES.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: OOCYTES AND IN SPECIALIZED CELL TYPES SUCH
 CC AS NEURAL RETINA CONES.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT OOGENESIS AND
 CC EMBRYOGENESIS.
 CC -1- DOMAIN: CONTAINS 37 FINGER MOTIFS IN 6 DOMAINS.
 CC -1- PTM: PHOSPHORYLATED.
 CC -1- SIMILARITY: BELONGS TO THE KRUPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: Contains 1 KRAB domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).

[1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE:88311086; PubMed:2842066;
 RA Morishita K., Parker D.S., Mucenski M.L., Jenkins N.A., Copeland N.G.,
 RT "Retroviral activation of a novel gene encoding a zinc finger protein
 in IL-3-dependent myeloid leukemia cell lines."
 RL Cell 54:831-840(1988).
 CC - SUBUNIT: MAY INTERACT WITH CTBP1 (BY SIMILARITY).
 CC - SUBCELLULAR LOCATION: Nuclear.
 CC - CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-11 IS THE INITIATOR.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M21829; AAA40581.1; ALT_INIT.
 DR PIR: A31591; A31591.
 DR HSSP: P08047; 1SP2.
 DR TRANSFAC: T00273; -.
 DR MGD; MGI:95457; Ev11.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00096; zf-C2H2; 10.
 DR ProDom; PDD00003; Znf_C2H2; 4.
 DR SMART; SM00355; Znf_C2H2; 10.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 10.
 KW Zinc-finger; Metal-binding; DNA-binding; Developmental protein;
 KW Repeat; Proto-oncogene; Chromosomal translocation.
 FT ZN_FING 21 44 C2H2-TYPE.
 FT ZN_FING 75 97 C2H2-TYPE.
 FT ZN_FING 103 125 C2H2-TYPE.
 FT ZN_FING 131 154 C2H2-TYPE.
 FT ZN_FING 160 182 C2H2-TYPE.
 FT ZN_FING 188 210 C2H2-TYPE.
 FT ZN_FING 217 239 C2H2-TYPE.
 FT ZN_FING 224 246 C2H2-TYPE.
 FT ZN_FING 752 775 C2H2-TYPE.
 FT ZN_FING 781 803 C2H2-TYPE.
 FT DOMAIN 421 434 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 877 928 ASP/GLU-RICH (ACIDIC).
 FT SITE 553 557 CTBP-BINDING MOTIF 1 (BY SIMILARITY).
 FT SITE 584 588 CTBP-BINDING MOTIF 2 (BY SIMILARITY).
 SQ SEQUENCE 1042 AA; 116847 MW; 8DEDF164F536D2FE CRC64;
 Query Match 6.9%; Score 345; DB 1; Length 1042;
 Best Local Similarity 19.9%; Pred. No. 2, 5e-09;
 Matches 212; Conservative 140; Mismatches 367; Indels 344; Gaps 47;

QY 240 VG-----LPVSHNGATKSIIDYLEKANE--AKACLOSLTDSRRQISNIKEKLEKT 290
 DB 353 VQDNKPVLELP-----ERSSEERPLEKISIDPQSSSLLDVSFTSGSGLFTTSGSDLES 415
 QY 291 LIDVTDKMKIENHSITPFGCFKESFPFGIPLHGHRYLCKNNEELKAVLQEHENI- 349
 DB 416 --DLESDEK-----CKEN-----GKFKDVSPLOINIA 442
 QY 350 -VFNKAGFVNDKALLSSVLESEKLTSPIN-PYKDHMSVLKAYYAMNNEPNSDELTKIS 407
 DB 443 SITNKE--HNHNSVPSASVEQSAVSGAVNDITAIASIAEKY-----GSTGVGLQ 494
 QY 408 I---AVGLPOEFYKWE--QRKYQYNSRSPSLERTSKPLAENSPTTQDLSLPRSP 461
 DB 495 DKVGVALPYPSMFLPFPFPAFSQMYPF-----PDRDLRSJPL-----KNEQSP 539
 QY 462 VKPMDSITSPSIALHNSVTSQDPLRLTKSHFTNITAVDKLHRSNTSPPLNLSST 521
 DB 540 -----SEVKKLQGSSESPEDLTTKRD 562
 QY 522 SKNHSSTYTPNSFSSSELOAPPLDLSLPOKREBPKGIIATKNKTKATSINDHNSVSS 581
 DB 563 EKPITSGPSKSGTPTATS-QDQPLDLSGSKR-----ASGTYLTPRKNHVGEEK 613
 QY 582 SENSD-EPLNLTFFIKERSNSNLDNKNNPVFGNPP---SAPLYTLPPLPQSAFPAPT 637
 DB 614 GSNMDTRF-----SSDGLQHARPTPF-MDPIYRVEKRLTLPDLAKK----- 657
 QY 638 FMPPIVQTSIPGLRPYGL---DQNSFLPMATVYTGATPADMOQRKRYORKQFGQDL 694
 DB 658 -----KYLRPSPGFLFHQMSAIEHMA-----EKLSFSA-L 688
 QY 695 LDGADVYSGLDMD---SDSCSRKKIKTESGMVACDLCDKTFOKSSLLRKYEHF 751
 DB 689 KPEASELLQVPSMSFPAPPTLPLNLRKKE-RITCRYCGKLFPSANLTRLRHT 747
 QY 752 GKRPHQCOICKAFKHKHLIEHSR-LHSGEKPYOCDGKGRFSGSGYQHMNRYSYC 810
 DB 748 GEQPRCKYCCORSFSSNLRHVANNHINKERPFCHLCDRFQGOQTNLDRL----- 800
 QY 811 KKEAEERAEAREAKGHLGPTELMNRAYQST-----TPQY-----SDSEER 856
 DB 801 -----KKHEHNMSTGATSSPHSELESAGAILDDKEDVYFEIRNFINSNHG 848
 QY 857 ESMRPGSEK--EHEKGEBSGYGLARRD--GDEEESE--ESESNNKMDTPE--- 906
 DB 849 SOSPRNMEERNMGSHFKD-KKALATQNSDLDDVEVEDEVLLDDEDDNDIDPKPRKEL 907
 QY 907 -TIRDEETGDHSMDSSEDEGKMETKSD-----HEEDNMEDQM 943
 DB 908 GYTRIDEELPE---DYREAGALEMSCKASPVRYKEEDYKGL 947
 RESULT 14
 Z208 HUMAN STANDARD; PRT; 1167 AA.
 ID 2208 HUMAN
 AC 043345;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein 208.
 GN ZNF208 OR ZNF91L.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE:98391764; PubMed:9724325;
 RA Eichler E.E., Hoffman S.M., Adamson A.A., Gordon L.A., McCreedy P.,
 RT "Complex beta-satellite repeat structures and the expansion of the
 zinc finger gene cluster in 19p12.";

RL Genome Res. 8:791-808(1998).
 CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- TISSUE SPECIFICITY: Ubiquitous.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.
 CC -1- SIMILARITY: Contains 1 KRAB domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: AC003973; AAB94784.1; --
 DR HSSP: P25490; IUBD.
 DR GeneW: HGNC:12999; ZNF208.
 DR MIM: 606760; --
 DR GO: GO:0005634; C:nucleus; NAS.
 DR GO: GO:0003677; F:DNA binding activity; NAS.
 DR GO: GO:0008270; F:zinc ion binding activity; NAS.
 DR GO: GO:0006355; F:regulation of transcription, DNA-dependent; NAS.
 DR InterPro: IPR001099; KRAB.
 DR InterPro: IPR007087; Znf_C2H2.
 DR Pfam: PF01352; KRAB; 1; C2H2.
 DR Pfam: PF00096; zf_C2H2; 33.
 DR ProDom: PD000003; Znf_C2H2; 16.
 DR SMART: SM00349; KRAB; 1.
 DR SMART: SM00355; Znf_C2H2; 33.
 DR PROSITE: PS50805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 33.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 34.
 DR Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 KM Nuclear protein; Repeat.
 FT 2N_FING 201 223 KRAB.
 FT 2N_FING 229 251 C2H2-TYPE (DEGENERATE).
 FT 2N_FING 257 279 C2H2-TYPE.
 FT 2N_FING 285 307 C2H2-TYPE.
 FT 2N_FING 313 335 C2H2-TYPE.
 FT 2N_FING 341 363 C2H2-TYPE.
 FT 2N_FING 369 391 C2H2-TYPE.
 FT 2N_FING 397 419 C2H2-TYPE.
 FT 2N_FING 425 447 C2H2-TYPE.
 FT 2N_FING 453 475 C2H2-TYPE.
 FT 2N_FING 481 503 C2H2-TYPE.
 FT 2N_FING 509 531 C2H2-TYPE.
 FT 2N_FING 537 559 C2H2-TYPE.
 FT 2N_FING 565 587 C2H2-TYPE.
 FT 2N_FING 593 615 C2H2-TYPE.
 FT 2N_FING 621 643 C2H2-TYPE.
 FT 2N_FING 649 671 C2H2-TYPE.
 FT 2N_FING 677 699 C2H2-TYPE.
 FT 2N_FING 705 727 C2H2-TYPE.
 FT 2N_FING 733 755 C2H2-TYPE.
 FT 2N_FING 761 783 C2H2-TYPE.
 FT 2N_FING 789 811 C2H2-TYPE.
 FT 2N_FING 817 839 C2H2-TYPE.
 FT 2N_FING 845 867 C2H2-TYPE.
 FT 2N_FING 873 895 C2H2-TYPE.
 FT 2N_FING 901 923 C2H2-TYPE.
 FT 2N_FING 929 951 C2H2-TYPE.
 FT 2N_FING 957 979 C2H2-TYPE.
 FT 2N_FING 985 1007 C2H2-TYPE.
 FT 2N_FING 1013 1035 C2H2-TYPE.
 FT 2N_FING 1041 1063 C2H2-TYPE.
 FT 2N_FING 1069 1091 C2H2-TYPE.
 FT 2N_FING 1097 1119 C2H2-TYPE.
 FT 2N_FING 1125 1147 C2H2-TYPE.
 SEQUENCE 1167 AA; 134352 MW; E2184DF23B0D35E9 CRC64;

Query Match 6.8%; Score 341.5; DB 1; Length 1167;
 Best Local Similarity 21.2%; Pred. No. 4,2e-09;
 Matches 190; Conservative 121; Mismatches 313; Indels 271; Gaps 40;
 QY 7 GNRKPCCTCGKAFKKYKHLKHLHSGEKYCEPCNCKRRFSHSGSVSHHSKKICIL 66
 DB 421 GKKPYKCECGKAFKFWSSNLMKHKIHTGETYKCECGKFSWSTLSYH----- 471
 QY 67 ISVNGMRNNIKTGSSP--NSVSSPTNSAITQLRNKLKNGKPLMSBOTG--LLKIKT 121
 DB 472 -----KKITVEKPYKCECGKAFNOSALLIKHKRIHTBEKPYKCECGKTSKYST 523
 QY 122 EPLDENDYVLMATNG--FSSGSPFNNGGLAT---SPGVHPSAQSPMQLGYCMELP 175
 DB 524 -----LTHKAIHAGEKPYKCECGKTFIVSTLTTHKAI---HAG--EKP 564
 QY 176 LIGFTPMNSNLSEVQVLOIVNTYSRQMDCKTEIDSLKGYHMKDPCSGQBEQVTV-- 233
 DB 565 Y-----KCECGKAFS-----KPSILYHKVYHGEKPYKCECGKAFN 603
 QY 234 -SPNIPVGLPVVSHNGATKSIIDYTLKVNBAKACLOSLTTSRQISNIKKELTLLI 292
 DB 604 MGSNL-----MEH---KRI--HTGEKPYKCECGKFSFVS----- 634
 QY 293 DLVTDKMIENHSISTPSCQPKESFPGPILHOHER-----YLCMNEBI----- 339
 DB 635 -VLTAKVLI--HTGEKPYKCECGKAYKWSSTLSYHKIHTVEKPYKC---ECCGKAFNR 688
 QY 340 KAVLOPHENIVNKNAGVPYDNKAL-----LLSVLSKGLTSPINPK----- 382
 DB 689 SALLIKHKRI-----HTBEKPYKCECGKTSKYSTLTTHKAIHAGEKPYKCECGKA 741
 QY 383 -DHMSYLKAYVAMNM--EPNSDELKLSIYAVGLPOFVKEWEPQKRYOYSNRSPSLER 439
 DB 742 FSKFSLTLTHKXVYHTEKPYKCE-----ECGKAYNM-----PSTLS 777
 QY 440 TSKPLAPNSNPTKOSLPRSPYKPMDSITSPSIAMENSVTSQDPLRLT--KSHT 496
 DB 778 YHKIHTGEKPYKCEEC-----GKGFSMFSILTKHVIHTGEKPYKCECGKAFSWL 829
 QY 497 NIKAVDKLHSRS-----NTSPMLNSTSSKNSHS-----SYT 531
 DB 830 SVFSKHKIHTAGEKPYKCEACGKAYVTSIL--TKHVYHTEKPYKCECGKAFNWS 885
 QY 532 PMSFSSEEL-----QAEPLD--LSLPKQMBEPKGIATKTKTKATSINLDHNSVSSS 582
 DB 886 SNLMEHKIHTGETPYKCECGKAFSWPSLSLTHKATTHAGEKPYKCECGKAFSWSRLL 945
 QY 583 EN-----SDEPLNLTPIKKEFSNSNNLDNKSNNPVFGMNPFAKPLVPLPQ----- 630
 DB 946 EHKATTHAGEBPYKCECGKAFNWSNL-----MEHKRIHTGEKPYKCECGK 992
 QY 631 --SAPRPATFMPVQVSIPLGLRPYGLDOMSLPHMAVYTPGAATFAMQGRK--YORK 687
 DB 993 SFTFSILTKHKVYHT--GEKPYK-CEECG---KAYVMSSTLSYHKIHTVEKPYKCE 1044
 QY 688 OGFGQGLDGAQDYMGDLMDTSDSLRKKIKTESGMVACDLCDTKFOKSSILRRK 747
 DB 1045 ECGKGYVM-----FSILAKHVYHTEKPYKCECGKAYKMPSTLRYHK 1088
 QY 748 YEHTGRPHQCOICCKAFKHKHLIHSRLHSGEKPYKCDCKGRFSHSGSVSQH 802
 DB 1089 KIHTEKPYKCECGKAFSTFSILTGHKVIHTEKPYKCECGKAFKAPSWLSVFSKH 1143
 RESULT 15
 ZF25 HUMAN STANDARD; PRT; 783 AA.
 ID ZF25 HUMAN
 AC 09UT15;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein ZFP25.

GN ZPD25.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20135598; PubMed=10673043;
RA Li X.-A., Kokame K., Okubo K., Shimokado K., Tsukamoto Y., Miyata T.,
Kato H., Yutani C.;
RT "Cloning and characterization of a novel gene encoding a zinc finger
protein with 25 fingers."
RL Biochim. Biophys. Acta 1489:405-412(1999).
CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- TISSUE SPECIFICITY: Expressed in brain, heart, skeletal muscle,
kidney and pancreas. Weakly expressed in aorta, liver and lung.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB027251; BA05623.1; -.
DR HSSP; P08047; 1SP2.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF00096; zf_C2H2; 19.
DR PRINTS; PR00048; ZINC_FINGER.
DR ProDom; PD00003; Znf_C2H2; 16.
DR SMART; SM00355; Znf_C2H2; 22.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 16.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 24.
KW Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;
KW Nuclear protein; Repeat.
FT ZN_FING 76 98 C2H2-TYPE (ATYPICAL).
FT ZN_FING 104 126 C2H2-TYPE.
FT ZN_FING 132 154 C2H2-TYPE.
FT ZN_FING 160 182 C2H2-TYPE.
FT ZN_FING 188 210 C2H2-TYPE.
FT ZN_FING 216 238 C2H2-TYPE (ATYPICAL).
FT ZN_FING 244 266 C2H2-TYPE.
FT ZN_FING 272 294 C2H2-TYPE (ATYPICAL).
FT ZN_FING 300 322 C2H2-TYPE (ATYPICAL).
FT ZN_FING 328 350 C2H2-TYPE.
FT ZN_FING 356 378 C2H2-TYPE.
FT ZN_FING 384 406 C2H2-TYPE (ATYPICAL).
FT ZN_FING 412 434 C2H2-TYPE.
FT ZN_FING 440 462 C2H2-TYPE.
FT ZN_FING 468 490 C2H2-TYPE.
FT ZN_FING 496 518 C2H2-TYPE.
FT ZN_FING 524 546 C2H2-TYPE (ATYPICAL).
FT ZN_FING 552 574 C2H2-TYPE (ATYPICAL).
FT ZN_FING 580 602 C2H2-TYPE (ATYPICAL).
FT ZN_FING 608 630 C2H2-TYPE.
FT ZN_FING 636 658 C2H2-TYPE.
FT ZN_FING 664 686 C2H2-TYPE.
FT ZN_FING 692 714 C2H2-TYPE.
FT ZN_FING 720 742 C2H2-TYPE.
FT ZN_FING 748 770 C2H2-TYPE (ATYPICAL).
SQ SEQUENCE 783 AA; 90672 MW; 84882575525DC34C CRC64;

Query Match 6.8%; Score 339.5; DB 1; Length 783;
Best Local Similarity 20.5%; Pred. No. 3.2e-09;
Matches 170; Conservative 94; Mismatches 304; Indels 263; Gaps 30;

11 FKTEGKAFKXKHLKHLRIHSGEKPEKPCNCKRPSHSGSYSHISSKXICGLISVN 70

Db 132 YKCECGAFWFTLTGKRIHTEGKPYKCECGKAFNQSOLTRH----- 178
Qy 71 GRMRNNITGSSPNVSSPTNSAITQRLNKLNGKPLSMWEGTGLKITEPLDFNDYK 130
Db 179 ---KIITEKPKCE-----ECGKAFKQASHLTIHKI----- 208
Qy 131 VLMATHGFSGSSPFNNGGIGATSPGVHPSAQSPPMHLGVGMWEPALGFPPTNSNLSVQ 190
Db 209 ---IH-TGKPYKCECGAFVFSQSHTLTK-ILHTGENL-----YKCECG 250
Qy 191 KVLQIVDVTVRQKMDCTEDI SKLKGTHMCDPSQPEEGVTSNPVGLPVYSHNGA 250
Db 251 KAFNLFNLTNHRHAGE-----KPYKCE-CGR-----AFNLSNINLK 289
Qy 251 TKSITDYLEKVNKAC-----LSLTDSRQISNIKE-----KLRTILDIVTDK 299
Db 290 QEKI--HTGKLNKCECDKAFNNSLKTAKKILMEKPYKCECGKAFNQSOLTRH 347
Qy 300 MIENHSISTPSCOPCKESFPGPIPLHQR-----YLCKMNEIKAVLQPHENIVPN 352
Db 348 II--HTGKPYKCECGAFNQSOLTRHKKIHTAKSYK-----ECGKAFNQHNLN 401
Qy 353 KAGVFD-----NKALLSYLS-EKGLTSPINDYKOHMSVLAAYANMEPNDE 402
Db 402 HRKIYSGKPYKCECGAFNRSSTLTNRKKIHTGKPYKCE-----ECDRAFQSQSLTE 457
Qy 403 LKISIANGLQEPFKWEPQGRKXYQYNSRSPSLERSKPLAPNSNTTDSILPRPV 462
Db 458 HKKI-----HTGKPYKCECGKAFNRSSTLTN----- 485
Qy 463 KPMDSITPSIAELNNSVSCDPLRLTKSSHFTIKAVDKLDHSRNTPSPLNLSSTSS 522
Db 486 -----HKRIHTEKPYKCECGKAFN-QSYQLTRHKIYHTKELNCKEERG 530
Qy 523 KNSHSSYTPNSPSSSEELQABPLDLSPKQMBEPKGIATYKNTKATYINLDNSVSSS 582
Db 531 KAFKSSH-----RTIHKIHTGKPYKCE-----EHGVFNQS 564
Qy 583 ENSDPLMLTPIKKEFSNNSNLDKSNPNVFGM--NPSAKPLVPLPQ-----S 631
Db 565 S-----NLTKQKIIHTEENLYKTEHGAFAFNLSNITNKKIYTGKPHKCECGKAYN 618
Qy 632 AAPPATFMPVOTSIPLGLRPYPGLDQMSFLPHMAVTPYGAATFADMOQRKYGQKGFQ 691
Db 619 RFSNLTTHKRIHT---GEKYQCAE-----CGKA----- 644
Qy 692 GDLLDGAODYMSGLDMDTSDSCLSRKKIKTESGMVACDLCDKTFQKSSILRHYEHT 751
Db 645 -----FNCSSTINRHKIHTGKPYKCECGKAFNLSSTLTAKKIIHT 687
Qy 752 GKRPQCOCCKKAFKXKHLKHLRIHSGEKPYQCDKGRKFSHSGSYSH 802
Db 688 GEKPYKCECGKAFNQSOLTRHKKIHTSEKPYKCECGKSFNPSINIH 738

Search completed: January 5, 2004, 16:29:00
Job time : 23 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2004, 16:27:10 ; Search time 46 Seconds
(without alignments)
5295.688 Million cell updates/sec

Title: US-09-964-238-2
Perfect score: 4991
Sequence: 1 MLTQAGNRKFKTECGAFKAF.....DGKMTKSDHEEDNEDMG 944

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: SP archaea:*
- 2: SP bacteria:*
- 3: SP fungi:*
- 4: SP human:*
- 5: SP invertebrate:*
- 6: SP mammal:*
- 7: SP mhc:*
- 8: SP organelle:*
- 9: SP phage:*
- 10: SP plant:*
- 11: SP rodent:*
- 12: SP virus:*
- 13: SP vertebrate:*
- 14: SP unclassified:*
- 15: SP viirus:*
- 16: SP bacteriap:*
- 17: SP archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4273	85.6	1214	13	Q91B60
2	4270	85.6	1213	13	Q91B61
3	1647.5	33.0	1104	4	Q8NB68
4	1646	33.0	306	11	Q8BSG9
5	1606.5	32.2	1154	4	Q13088
6	1449.5	29.0	1078	13	Q9DEH7
7	817.5	16.4	747	5	Q9VA39
8	817.5	16.4	1054	5	Q9VA40
9	757	15.2	650	5	Q94196
10	533.5	10.7	589	5	Q8MSQ8
11	365	7.3	189	11	Q8CSP2
12	343.5	6.9	914	11	Q8C827
13	341.5	6.8	949	11	Q35700
14	331.5	6.6	743	11	Q8R5D1
15	331.5	6.6	751	4	Q8TBA9
16	331.5	6.6	752	11	Q8BT00

17	328.5	6.6	569	4	Q9H9S6	Q9H9S6 homo sapien
18	328.5	6.6	701	4	Q9UG02	Q9UG02 homo sapien
19	324.5	6.5	819	4	Q8N946	Q8N946 homo sapien
20	322.5	6.5	688	4	Q8TA05	Q8TA05 homo sapien
21	321.5	6.4	616	4	Q8N9M3	Q8N9M3 homo sapien
22	319.5	6.4	1031	11	Q8CJ78	Q8CJ78 mus musculu
23	315	6.3	578	11	Q8BLA2	Q8BLA2 mus musculu
24	313.5	6.3	648	11	Q8K167	Q8K167 mus musculu
25	312.5	6.3	599	4	Q8TF20	Q8TF20 mus sapien
26	309	6.2	898	13	Q12958	Q12958 oryzias lat
27	306.5	6.1	1261	13	Q90ZM5	Q90ZM5 gallus galli
28	306	6.1	625	11	Q8BIN6	Q8BIN6 mus musculu
29	304.5	6.1	623	11	Q62510	Q62510 mus musculu
30	302.5	6.1	821	4	Q96JF6	Q96JF6 homo sapien
31	302.5	6.1	1173	13	Q90783	Q90783 gallus galli
32	299.5	6.0	1615	13	Q57415	Q57415 gallus galli
33	298.5	6.0	1186	11	Q08961	Q08961 rattus norv
34	297	6.0	468	4	Q8TF47	Q8TF47 homo sapien
35	297	6.0	738	4	Q9ULAI	Q9ULAI homo sapien
36	297	6.0	871	4	Q8WTR7	Q8WTR7 homo sapien
37	296.5	5.9	637	4	Q96IR2	Q96IR2 homo sapien
38	296.5	5.9	1377	13	Q9DDN5	Q9DDN5 xenopus lae
39	296	5.9	786	4	Q8TDG8	Q8TDG8 homo sapien
40	295.5	5.9	769	4	Q81ZD3	Q81ZD3 homo sapien
41	295	5.9	648	4	Q81YB9	Q81YB9 homo sapien
42	294	5.9	914	4	Q9ULS9	Q9ULS9 homo sapien
43	292.5	5.8	714	4	Q8TE95	Q8TE95 homo sapien
44	291	5.8	873	6	Q9XSR1	Q9XSR1 canis famli
45	290.5	5.8	1350	13	Q91929	Q91929 xenopus lae

ALIGNMENTS

RESULT 1	Q91B60	PRELIMINARY;	PRT;	1214 AA.
ID	Q91B60			
AC	Q91B60;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DR	Smad interaction protein 1.			
GN	SIPL.			
OS	Xenopus laevis (African clawed frog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;			
OC	Xenopodinae; Xenopus.			
OX	NCBI_TaxID=8355;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Eisaki A., Kuroda H., Hayata T., Asashima M.;			
RT	"Xenopus laevis SIPL.";			
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AB038353; BAA94081.1; -			
DR	InterPro; IPR001356; Homeobox.			
DR	InterPro; IPR001087; Znf_C2H2.			
DR	Pfam; PF00096; ZF-C2H2; 8.			
DR	ProDom; PD000003; Znf_C2H2; 1.			
DR	SMART; SM00389; HOX; 1.			
DR	SMART; SM00355; Znf_C2H2; 7.			
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.			
DR	PROSITE; PS00157; ZINC_FINGER_C2H2_2; 6.			
KW	Metel-Pindng; Zinc; Zinc-finger.			
SQ	SEQUENCE 1214 AA; 136872 MW; 8D64868F652FB325 CRC64;			

Query Match 85.6%; Score 4273; DB 13; Length 1214;
Best Local Similarity 86.1%; Pred. No. 3.6e-256;
Matches 815; Conservative 56; Mismatches 68; Indels 8; Gaps 5;

QY	1	MLTQAGNRKFKTECGAFKFKHLKHLRIHSGEKYECPNCKKRSKSGSYSHSS 60
DB	272	MLTQAGNRKFKTECGAFKFKHLKHLRIHSGEKYECPNCKKRSKSGSYSHSS 331

```

QY 61 KCCIGLISVNGRNNIKTGSSPNSVSSPTNSAITQLRNLKENGKPLSMSEQTGLKIK 120
DB 332 KCCIGLISVNGRNNIKTGSSPNSVSSPTNSAITQLRNLKENGKPLSMSEQTGLKIK 391
QY 121 TEBLDNDYKVLMAHNGSSGSPFNNGIGATSPGCHPSAOSPMOHGVCMEAPLGF 179
DB 392 TEBLDNDYKVLMAHNGSSGSPFNNGIGATSPGCHPSAOSPMOHGVCMEAPLGF 451
QY 180 PTMNSNLSEVQKVLQIVDNTVSRQKMDCTEDISKLGVMKQPCSOPEOGVTSFNIP 239
DB 452 PTMNSNLSEVQKVLQIVDNTVSRQKMDCTEDISKLGVMKQPCSOPEOGVTSFNIP 511
QY 240 VGLPVVSHNGATKSIIDYTLKENVAKACLSITDSRQISNIKKKRLTLDLVTDK 299
DB 512 VGLPVVSHNGATKSIIDYTLKENVAKACLSITDSRQISNIKKKRLTLDLVTDK 571
QY 300 MLENHSITSPFGQFCKESFPGPIPLHQRHRYLCRNKEIKAVLOPHENIVPKAVFPD 359
DB 572 MLENHSITSPFGQFCKESFPGPIPLHQRHRYLCRNKEIKAVLOPHENIVPKAVFPD 631
QY 360 NKALLLSVLSKGLTSPINPYKDHSVLYKAYYAMNENPBDLKTSLAVGLPOEFVKE 419
DB 632 KQALLLSVLSKGLTSPINPYKDHSVLYKAYYAMNENPBDLKTSLAVGLPOEFVKE 691
QY 420 WFOQRVYQYSNRSRPSLERTS--KPLAPNSPTTKDLSLPSVPKMDSITSPSIAELH 477
DB 692 WFOQRVYQYSNRSRPSLERTS--KPLAPNSPTTKDLSLPSVPKMDSITSPSIAELH 749
QY 478 NSTVSCDPLRLTKSHFTNIKAV-DKLDHSRNTSPPLNLSSTSKNSHSSSTPNSFS 536
DB 750 NRSVNCDTPLRLTKSHFTNIKAV-DKLDHSRNTSPPLNLSSTSKNSHSSSTPNSFS 809
QY 537 SEELQAEPLDLSPKQMRBEPKGIATKNTKATSIINDHNSVSSSENSEPLNTFFIKK 596
DB 810 SEELQAEPLDLSPKQMRBEPKGIATKNTKATSIINDHNSVSSSENSEPLNTFFIKK 869
QY 597 EFSNNSNLDKSNPNVFGANPFSAPKLYTLPQSAFPATFMPVQTSIPGLRPYGLD 656
DB 870 EFSNNSNLDKSNPNVFGANPFSAPKLYTLPQSAFPATFMPVQTSIPGLRPYGLD 927
QY 657 QMSFLPHMATYTPGATFADMOORRYQKQGFQDLDGADYMSGLDMDTDSCLS 716
DB 928 QMSFLPHMATYTPGATFADMOORRYQKQGFQDLDGADYMSGLDMDTDSCLS 987
QY 717 RKIKIKTESGMVACDLCTFKOKSSILRKYHTGKRPQCOICCKAFGKHLLIHSR 776
DB 988 RKIKIKTESGMVACDLCTFKOKSSILRKYHTGKRPQCOICCKAFGKHLLIHSR 1047
QY 777 LHSGEKPYQCDKCGKRFSGSYSGHMHRYSCKRAEEREAERAREKHLGPTELL 836
DB 1048 LHSGEKPYQCDKCGKRFSGSYSGHMHRYSCKRAEEREAERAREKHLGPTELL 1107
QY 837 MNNAVYQSTTPQGSDBEREEMPRDGRLEHEKEEDDYYDKLRQVGEFEFEFEES 896
DB 1108 MNNAVYQSTTPQGSDBEREEMPRDGRLEHEKEEDDYYDKLRQVGEFEFEFEES 1167
QY 897 ENKSMPTPETIRDEETGDSHMDSDSEDMETKSHHEEDNMDGM 943
DB 1168 ENKSMPTPETIRDEETGDSHMDSDSEDMETKSHHEEDNMDGM 1214

```

```

RESULT 2
Q918P1 PRELIMINARY; PRT; 1213 AA.
AC Q918P1;
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DE 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Sma-d-interacting protein 1.
GN S1P1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

```

```

OC Xenopodinae; Xenopus.
OX NCBI_TaxId=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20302561; PubMed=10842070;
RA van Grunsven L.A., Papin C., Avalosse B., Opdecamp K., Huybregt D.,
RT Smith J., Bellefroid E.;
RA "XSP1, a Xenopus zinc finger/homeodomain-encoding gene highly
RT expressed during early neural development.";
RL Mech. Dev. 94:189-193 (2000).
DR EMBL; AF237679; AAF81689.1; -.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 8.
DR Prodom; PD000003; Znf_C2H2; 1.
DR SMART; SM00355; ZNF_C2H2; 7.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 6.
DR Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 1213 AA; 136775 MW; E0AB78D294A2877B CRC64;

Query Match 85.6%; Score 4270; DB 13; Length 1213;
Best Local Similarity 86.0%; Pred. No. 5,6e-256;
Matches 814; Conservative 57; Mismatches 68; Indels 8; Gaps 5;

QY 1 MTQAGNKKFKTECGKAFKTKHLKEHLRHSGEKPYECNCKRFSGSHSS 60
DB 271 MTQAGNKKFKTECGKAFKTKHLKEHLRHSGEKPYECNCKRFSGSHSS 330
QY 61 KCCIGLISVNGRNNIKTGSSPNSVSSPTNSAITQLRNLKENGKPLSMSEQTGLKIK 120
DB 331 KCCIGLISVNGRNNIKTGSSPNSVSSPTNSAITQLRNLKENGKPLSMSEQTGLKIK 390
QY 121 TEBLDNDYKVLMAHNGSSGSPFNNGIGATSPGCHPSAOSPMOHGVCMEAPLGF 179
DB 392 TEBLDNDYKVLMAHNGSSGSPFNNGIGATSPGCHPSAOSPMOHGVCMEAPLGF 451
QY 180 PTMNSNLSEVQKVLQIVDNTVSRQKMDCTEDISKLGVMKQPCSOPEOGVTSFNIP 239
DB 452 PTMNSNLSEVQKVLQIVDNTVSRQKMDCTEDISKLGVMKQPCSOPEOGVTSFNIP 511
QY 240 VGLPVVSHNGATKSIIDYTLKENVAKACLSITDSRQISNIKKKRLTLDLVTDK 299
DB 512 VGLPVVSHNGATKSIIDYTLKENVAKACLSITDSRQISNIKKKRLTLDLVTDK 571
QY 300 MLENHSITSPFGQFCKESFPGPIPLHQRHRYLCRNKEIKAVLOPHENIVPKAVFPD 359
DB 572 MLENHSITSPFGQFCKESFPGPIPLHQRHRYLCRNKEIKAVLOPHENIVPKAVFPD 631
QY 360 NKALLLSVLSKGLTSPINPYKDHSVLYKAYYAMNENPBDLKTSLAVGLPOEFVKE 419
DB 632 KQALLLSVLSKGLTSPINPYKDHSVLYKAYYAMNENPBDLKTSLAVGLPOEFVKE 691
QY 420 WFOQRVYQYSNRSRPSLERTS--LAPNSPTTKDLSLPSVPKMDSITSPSIAELH 477
DB 692 WFOQRVYQYSNRSRPSLERTS--LAPNSPTTKDLSLPSVPKMDSITSPSIAELH 749
QY 478 NSTVSCDPLRLTKSHFTNIKAV-DKLDHSRNTSPPLNLSSTSKNSHSSSTPNSFS 536
DB 750 NRSVNCDTPLRLTKSHFTNIKAV-DKLDHSRNTSPPLNLSSTSKNSHSSSTPNSFS 809
QY 537 SEELQAEPLDLSPKQMRBEPKGIATKNTKATSIINDHNSVSSSENSEPLNTFFIKK 596
DB 810 SEELQAEPLDLSPKQMRBEPKGIATKNTKATSIINDHNSVSSSENSEPLNTFFIKK 869
QY 597 EFSNNSNLDKSNPNVFGANPFSAPKLYTLPQSAFPATFMPVQTSIPGLRPYGLD 656
DB 870 EFSNNSNLDKSNPNVFGANPFSAPKLYTLPQSAFPATFMPVQTSIPGLRPYGLD 926
QY 657 QMSFLPHMATYTPGATFADMOORRYQKQGFQDLDGADYMSGLDMDTDSCLS 716
DB 928 QMSFLPHMATYTPGATFADMOORRYQKQGFQDLDGADYMSGLDMDTDSCLS 986

```

QY 717 RKIKKTESGWACDLCKTFOKSSSLRHKYEHTGKSPHOCICCKAFKHHHLEHSR 776
 DB 967 RKIKKTESGWACDLCKTFOKSSSLRHKYEHTGKSPHOCICCKAFKHHHLEHSR 1046
 QY 777 LHSGEKPYQCDCKGRFSGSYSGYSHMNRYSYCKREAEERAEAREAREKCHLEPTLL 836
 DB 1047 LHSGEKPYQCDCKGRFSGSYSGYSHMNRYSYCKREAEERAEAREAREKCHLEPTLL 1106
 QY 837 MNRAVLOSITPGYSDSERESMPRDSSEKHEKGEKGYGKLRRPRGDEEEEEEES 896
 DB 1107 MNRAVLOSITPGYSDSERESMPRDSSEKHEKGEKGYGKLRRPRGDEEEEEEES 1166
 QY 897 ENKSMYDTPETIRDEEETGDHSMDDSDSKMETSKSDHEEDNMDQM 943
 DB 1167 ENKSMYDTPETIRDEEETGDHSMDDSDSKMETSKSDHEEDNMDQM 1213

RESULT 3

Q8NB68 PRELIMINARY: PRT: 1104 AA.

AC Q8NB68: 01-OCT-2002 (TREMblrel. 22, Created)
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Hypothetical protein FLJ34159.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Tissue=Brain;
 RA Ieshbahi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
 RA Houta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
 RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
 RA Ishi S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
 RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
 RA Megatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Negahari K., Masuno Y., Nagai K., Isogai T.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK091478; BAC03673.1; -;
 DR InterPro: IPR001356; Homeobox.
 DR InterPro: IPR007087; Znf_C2H2.
 DR Pfam: PF00096; Zf_C2H2; 3.
 DR ProDom: PD00010; Homeobox; 1.
 DR ProDom: PD00003; Znf_C2H2; 2.
 DR SMART: SM00389; HOX; 1.
 DR SMART: SM00355; Znf_C2H2; 7.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 6.
 KW Hypothetical protein; Metal-binding; zinc; zinc-finger.
 SQ SEQUENCE 1104 AA; 121919 MW; E0F3C8E2D3A1CFD CRC64;

Query Match 33.0%; Score 1647.5; DB 4; Length 1104;
 Best Local Similarity 41.8%; Pred.No.14e-93;
 Matches 405; Conservative 146; Mismatches 310; Indels 107; Gaps 32;

DB 386 -----SINISDIQNVLVKVAADVIRQVLE---NNQANLASKQDETINASPIQGGSHV 436
 QY 236 NIPPVGLPVYHNKATSIIDYTLKVNKACQSLTTSRQROSNKKKLTLLDLY 295
 DB 437 -ISATSLPVQDGTTLIIVSLQPSQLOVDPNKKENPVATNSCKSEKLPEDLVK 495
 QY 296 TD-DRMIEBH-SISTPFCQCKESFPGP-PLHQHERYLCMNEEIKAVLQPHENIVPN 352
 DB 496 SEKXSPFGVNDSTCLDCDC---PDIALBELKHY-----DLQPPQ-----PP 540
 QY 353 KAGVFDNKLALLSSVSEKLTSPINPYKDHMSVLKAYVAMNEPNSDELKISIAVL 412
 DB 541 LPAEAERDESSVSATDGNLSPSQPLKWLISLKAYVALNAPSEELSLADSVNL 600
 QY 413 POFVKEFERQKRYQVYNSNSPSLERKSLAPLPSNNTTDSILPRPVKPMDSITPS 472
 DB 601 PLDVVKKMFERMQAGQIS-----VQSESPSPPEPKV-----NIPAKNDQPOGAN 646
 QY 473 IAEILNSVTSQDPLRLTSSHTNIKAVDKLDHSRNTPEPLNLSSTSKSHSSSYTP 532
 DB 647 ANEPQDSTVNLQSLPKMTNS---PVLPGVSTTNGSRSTPSPPLNLSSTANTQGYLYTA 703
 QY 533 NSFSSSEIQAEPLDLSLPKQREPKGIATNKTVAISINDHNSVSSSSNS-----DE 587
 DB 704 EG-AQEBQVPEPLDLSLPKQGE-----LLERSTTVYQNSVYVQEE 746
 QY 588 PLNLFTIKERSNSNNLDKNNPNVFGNPPSAKPLVPLPQGAFFPATMPQVSTIP 647
 DB 747 PLNLSCKAEKQKQKSCVTD--SEPVNVIPPSANFINAIPVYAQLPTIVAIDQNSVP 804
 QY 648 GLRPPGLDQMSFLPHMAVTPYGAATPADMQORRYOKQKQFQGDLLDGAQDVMSGDD 707
 DB 805 CLRALAANKQITLTPQAVYITSTVSP-AVQEPPLKVIQPNQNDERDSTSEGSNVED 863
 QY 708 MTDSQSLRRKIKKTESGWACDLCKTFOKSSSLRHKYEHTGKSPHOCICCKAFK 767
 DB 864 QNDSSTPPKKMKRTENGWACDLCKTFOKSSSLRHKYEHTGKSPHOCICCKAFK 923
 QY 768 KHHLEHRLHSGEKPYPQCDCKGRFSGSYSGYSHMNRYSYCKREAEERAEAREK 827
 DB 924 KHHLEHRLHSGEKPYPQCDCKGRFSGSYSGYSHMNRYSYCKREAEERAEAREK 979
 QY 828 GHLGPTLLMRAVLOSITP-QGYSDSERESMPR--DGESEKHEKGE--GYNK 879
 DB 960 -EAGP-EILSEHYGARASPSQG--DSDERESLTREEDDESEKHEKHELOEKE 1035
 QY 880 LRRRDGEDEEESEENKSMYDTPETIRDEEETGDHSMDDSD-----GKMET 930
 DB 1036 CEKPGDEDEEBEVEEVEEVE-EAENGEAKTEGLMKODRAESQASISGQKVGESSE 1094
 QY 931 KSDHEEDN 938
 DB 1095 QVSEKTN 1102

RESULT 4

Q8BSG9 PRELIMINARY: PRT: 306 AA.

AC Q8BSG9: 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Zinc finger homeobox 1b.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Mesonephros;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium;

RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK032970; BAC28104.1; -
SQ SEQUENCE 306 AA; 35506 MW; 730827A3D350464C CRC64;

Query Match 33.0%; Score 1646; DB 11; Length 306;
Best Local Similarity 99.7%; Pred. No. 3.2e-94;
Matches 304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 639 MPVQVSIPLRLRYPLGLDQMSLPHMAVYTPGAATFADMQQRKTKYQKQGFQGDLLDGA 638
DB 1 MPVQVSIPLRLRYPLGLDQMSLPHMAVYTPGAATFADMQQRKTKYQKQGFQGDLLDGA 60
QY 699 ODYMSGLDMDTSDSCSLRKKIKKTESGMYACDLCKTFQKSSSLRHKYEHTGKPHQC 758
DB 61 QDYMSGLDMDTSDSCSLRKKIKKTESGMYACDLCKTFQKSSSLRHKYEHTGKPHQC 120
QY 759 QICKKAFKXKHLIEHSRLHSGEKPYQCDKCGKRFSGSGYQHMNRYSYCKREAEERE 818
DB 121 QICKKAFKXKHLIEHSRLHSGEKPYQCDKCGKRFSGSGYQHMNRYSYCKREAEERE 180
QY 819 AAREAREKCHLGTPLLMRAVLOSTTPQGYSDSEEREMPRDGSSEKHEKEGEGYG 878
DB 181 AAREAREKCHLGTPLLMRAVLOSTTPQGYSDSEEREMPRDGSSEKHEKEGEGYG 240
QY 879 KLRRRDGEDEEESEENKSMOTPETIRDEEETGDSMDSDSEDGKMETSDHEEDN 938
DB 241 KLRRRDGEDEEESEENKSMOTPETIRDEEETGDSMDSDSEDGKMETSDHEEDN 300
QY 939 MEDGM 943
DB 301 MEDGM 305

RESULT 5
Q13088 PRELIMINARY; PRT; 1154 AA.

AC Q13088;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE ZEB (Fragment).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Genetica T., Ruzizinsky D., Kadesch T.;
RT "Displacement of an E-box-binding repressor by basic-helix-loop-helix
RT proteins: Implications for B-cell specificity of the immunoglobulin
RT heavy-chain enhancer."
RL Mol. Cell. Biol. 15:0-0(1995).
DR EMBL; U19969; AAA62155.1; -
DR TRANSFAC; T00625; -
DR TRANSFAC; T02324; -
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR007087; Znf C2H2.
DR Pfam; PF00096; Zf-C2H2; 7.
DR ProDom; PD000010; Homeobox; 1.
DR ProDom; PD000003; Znf C2H2; 2.
DR SMART; SM00389; Hox; 1.
DR SMART; SM00355; Znf C2H2; 7.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 6.
KW Metal-binding; Zinc; Zinc-finger.
FT NON TER 1
SQ SEQUENCE 1154 AA; 126924 MW; 1A4DAC7851909113 CRC64;

Query Match 32.2%; Score 1606.5; DB 4; Length 1154;
Best Local Similarity 41.1%; Pred. No. 5.2e-91;

Matches 399; Conservative 142; Mismatches 314; Indels 115; Gaps 32;

QY 2 LTQAGNRKFKTECGKAFKYGKHLKHLRLHSGEKPYECPCKRFSHSGSYSHSSK 61
DB 265 VTQSGNRKFKTECGKAFKYGKHLKHLRLHSGEKPYECPCKRFSHSGSYSHSSK 324
QY 62 KICGLISVNGRMNNTKG--SSPNSVSSPTNSAITQLRNLKENGKPLSMBOQLKXI 119
DB 325 KICSLIPVNGRPRTKISQCSSP-SLSASPSPPRPQIRKXEN-KPL--OEOISLVNOI 380
QY 120 KTEPLDFNDYKVLMTHTGSSGSPFNNGGLATSPPLGVHPASQSPQOH---GVGAEPL 176
DB 381 KTEPLVDY-EFKPLVVASGINCSTPLQGVFTGGGFLQATSSPQMVQAVVLPVGLVSP 439
QY 177 LGEPITNSMISEVQKLTQ-VDNVTSRQMDKTEDISKLGKYNMKPCSOPEQGVTSR 235
DB 440 -----SILSLDIQNLKAVADGNVIRYLE---NNQANLAKGEQETINASPIQQGHGV 490
QY 236 NIPVGLPVYSHNGATKSIIDYTLKVNFAKACLOSLTDSRQISNIKKELRTIDLV 295
DB 491 -LSAISLPLVDQDTTKIINYSLEQPSQLQVVPQNLKKNFVATNSCKSEKLPEDLVTK 549
QY 296 TD-DKMIENH-SISPPSCQPKESPPI-PLHNERVLCQNNERIKAVLPHEHIVN 352
DB 550 SEKDYSFEGGVNDSTCLCDDC---PGDINALPELKH-----DLKOPTP-----PP 594
QY 353 KAGVPVQNKALLLSVLSSEKGLTSPINPKYDHSVYLKAYAMNMBNSDELKISAVGL 412
DB 595 LPAAEAKREKSSVSAITGNGINLSPSPQPLKNLISLKAYVALNAQSAEELKINDSVNL 654
QY 413 PQEFVKEPFQKRVYQVYNSRSPSLERTSKPLAPNSNPTTKOSLLPRGVKPMDSITSPS 472
DB 655 PLDVVKMFEPKQACQIS-----VQSESEPSPEPGKV-----NIPAKNNDQPOSAN 700
QY 473 IAEIHNSVTSQCPRLTSSHTNKAVDKLDHSNSNTPPLNLSSTSSKSHSSSYRP 532
DB 701 ANEPQDSTVNLQSLPKMTNS---PVLPGVSTTNGSRSTSPSPPLNLSSRNTQGYLYTA 757
QY 533 NSFSSEILOAEPDLISLPKQRE-----PKGIATKTKTATSIINLHNSVSSSENS 585
DB 758 EG-AQEEPPQVPELDLSLPQGGELLERILPLPQCLFCPRT----- 797
QY 586 DEPLNTITFKKPSNNSNIDNKSNNPFQGMPPSAKPLTPLPPOSAPFPATFMPVQTS 645
DB 798 ---LELVLAKEEPQKSCVTD--SEPVVAVIPPSANPINIAIPVYAQLPTVAIDQNS 852
QY 646 IGLRYPPLDQMSFLPHMAVYTPGAATFADMQQRKTKYQKQGFQGDLLDGAQDVMGSL 705
DB 853 VPCLRALANKOTILIPVAYTISTVSP-AVQEPPLKVIQPNGNODERQDTSBGVSNV 911
QY 706 DDMTSDSCSLRKKIKKTESGMYACDLCKTFQKSSSLRHKYEHTGKPHQCQICKKAF 765
DB 912 EONDSDSIFPKKKRKTENGMYACDLCKITQKSSSLRHKYEHTGKPHCGICKKAF 971
QY 766 KXKHLIEHSRLHSGEKPYQCDKCGKRFSGSGYQHMNRYSYCKREAEERPAAREAR 825
DB 972 KXKHLIEHSRLHSGEKPYQCDKCGKRFSGSSSSQHMNRYSYCKREAEERDSTQOE-- 1029
QY 826 EKCHLGTPLLMRAVLOSTTP-QGYSDSEEREMSR--DGSSEKHEKEGE-----GY 877
DB 1030 ---EAGP-BILSNHVGARAPSPQG--DSDEBSLTFREDEDESEKEEEDKEMELOEE 1083
QY 878 GKLRRRDGEDEEESEENKSMOTPETIRDEEETGDSMDSDSED-----GKM 928
DB 1084 KECEKQGGEBEEREBEVEEVE-FAINBGEAKTEBLMDRAEASASLGQKVGSS 1142
QY 929 ETKSDHEEDN 938
DB 1143 SEQVSEKTN 1152

RESULT 6
Q9DEH7

ID Q9DEH7 PRELIMINARY; PRT; 1078 AA.
AC Q9DEH7;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Kheper.
GN ZFH1 OR KHEPER.
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_Taxid=7955;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20541661; PubMed=11087624;
RA Murooka O., Ichikawa H., Shi H., Okumura S., Taira E., Higuchi H.,
RA Hirano T., Hibi M., Miki N.;
RT "Kheper, a Novel ZFH/deltaEF1 Family Member, Regulates the Development
of the Neuroectoderm of Zebrafish (Danio rerio).";
RL Dev. Biol. 228:29-40(2000).
DR EMBL; AB016799; BAB18865.1; -.
DR ZFIN; ZDB-GENE-010621-1; zfhx1.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zfc_C2H2; 1.
DR Pfam; PF00096; zfc_C2H2; 7.
DR Prodom; PD000010; Homeobox; 1.
DR Prodom; PD000003; Znf_C2H2; 2.
DR SMART; SM00389; Hox; 1.
DR SMART; SM00355; Znf_C2H2; 7.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 7.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 1078 AA; 117844 MW; B3D236A9D1235BC CRC64;

Query Match 29.0%; Score 1449.5; DB 13; Length 1078;
Best Local Similarity 37.6%; Pred. No. 2.6e-81;
Matches 376; Conservative 126; Mismatches 285; Indels 213; Gaps 31;

QY 5 GAGNRKFKCTEGCKAFKTKHKLHRLHSGEKPYECNCKRFSHSGSYSHSKCI 64
DB 203 GGGNKKFKCTEGCKAFKTKHKLHRLHSGEKPYECNCKRFSHSGSYSHSKCI 262
QY 65 GLISVNGMRNNIKTG-----SSPNVSSSPTSATQLNKKENGKPLMSQTGL 117
DB 263 GLISVNGMRNNIKTG-----SSPNVSSSPTSATQLNKKENGKPLMSQTGL 117
QY 118 KIKTEPLDFNDYKVMATHTGFSGSPFNNGIGATSPGVHPSAOSPMOHLGVMEAPL 177
DB 319 QIKSEPLDY-EYKPVVVAPSAAGVNGMPQG-----AAPLQAGVAVVLPV 365
QY 178 GEPFT-MANSLSVQYLOI-VDNTVSRQMDCKTEDISKLYHMKDPCSGPEEGV--- 232
DB 366 GLVSPISINLGLQNVLVKAVDGVNIRQVLBS-----TQAK-----QAGTGIYGA 412
QY 233 -----TSNTPPVGLPVVSHNGATKSIIDYTLKVNKAKALQSLTDSRRQINIKKE 286
DB 413 GGIAATQOQVATLPIIDQGNNAKIFVNSLDP-SQQAVALQS-----PKKE 460
QY 287 KLRITLIDVTDKMLTNSHISTPSCOFCKESPFGPIPIHQHERYLCKNNEIKAVLOPH 346
DB 461 PLGSTVTCCKGKGLPDLTVKT-----NDR----- 485
QY 347 ENIVNKAQVFDNKAALISVLS-----KGLTSPINPYKDHMS 386
DB 486 -----KTLTYDEKSMANLILKRCGKDEHRINGKNLEKMDLEGLCPCGPPLKNLIS 539
QY 387 VLKAYVAMNENSELKISIAVGLPOEFVKEWFEORKAVOYS-NSRSPSLERTSKPLA 445
DB 540 LKAFALANNEPTKELAKISVSLSIPAVVKKWFKEMQLOIGISNDPSSPQHEE----- 593
QY 446 PMSNPTTQDILPRSPVVKWMSITSPSIAELHNSVTSQDPLRLTKYSHFTNIKAVDKLD 505

DB 594 EOTTPEVDLDGTRKASPKEDLDEQWN---SEKOEERCCSPARGIAAS-----VNGIESVP 645
QY 506 HSRNSPTPLNLSSSSSKNSHSSSYTPNSFSSEELQAPLPLSLDKORREPKGIATNKK 565
DB 646 AS---PSPNLNLSAG-----PVPARTLEEGEGLDLSLP-----KSA 679
QY 566 TKATISINLDHNSVSSSSSENSDEPLNLPFKKEFSNNTLDNKNPVGCMNPFSAKP--- 622
DB 680 TTASVASHANTVYSAGQ---EPNLITCTKEL-----LSNASSTAAIAYASPSANPINI 731
QY 623 LYTPLPPOSAPPATMPPVOTSTIPGLRPYPGLDQMSFLPHMAYTYPGATFADMOOR 682
DB 732 MTTQLPPLVATD-----QGAQCLRALTTTKQTILIPOLTVSYTTTSSSPAGTDTPQ 784
QY 683 K-YRKQSGFGDLDLDGADVWMSGUDMDPDSQCSRRKIKKTESGMVACDICTKFOKSS 741
DB 785 KNILHVNQIKKEKQDMSGSEATSLTEQDSDSGPKKKMKRTESGLVACDICTKFOKSS 844
QY 742 SLNRHRYHTGKRPHQCOICKAFKHKHLIEHSLHSGEKPYQCDCKGKRFSGYSYQ 801
DB 845 SLNRHRYHTGKRPHECICSKAFKHKHLIEHSLHSGEKPYQCDCKGKRFSGYSYQ 904
QY 802 HMRHRYSTCKRBAER-----EAEERAREKHLGPTLLMTRAYLOSITPGYSDBE 855
DB 905 HMRHRYSTCKRBAER-----EAEERAREKHLGPTLLMTRAYLOSITPGYSDBE 954
QY 856 RESMPRDESEKHE-----KEGSEYCKLARRDGEDEEESESEENKS 900
DB 955 RGSTRDSESEDEBMGSLVDEDEIQVVKIGEDBDQADBDQERMBEDGDEBGBG 1014
QY 901 MDTPETIRDEETGDSHSDSDSEDKMETKSDHEEDNME 940
DB 1015 ETLKVVWQDCBDEBNEDQTSBDADT-SDHTEEMTE 1053

RESULT 7
Q9VA39 PRELIMINARY; PRT; 747 AA.
ID Q9VA39;
AC Q9VA39;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE ZFH1 protein.
GN ZFH1 OR CG1322.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Mecoptera; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Hoskins R.A., Galle R.F.,
RA Amanatides P.G., Scherer S.B., Li P.A., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Suton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blaise R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle J., Baxter E.G., Helt G., Nelson C.R., McKlos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferreira C., Ferriere S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan F., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,


```

QY 129 YKVLMTAGFSGSSPPMNGGLATSPGLVHPAQAQSPMOHLGVGMEAPLLGF-PTMNSLNS 187
DB 91 YOLMM-----YRNMILQLOTPQVSLPSTANNM 119
QY 188 EVQKVIQIDNVYSRQKMDCKTEDISKLGYNHKDCSQPBEQGVTSNIPPIGLPVYSH 247
DB 120 DVMSLQ-----ANLFOSLNGTSPTTQESAPASPEKLEVVDEEVS- 164
QY 248 NGATKSIIDYLEKVNKAACLOSLTTDSRQISNIKEKRTLIDLVTDKMIENHSIS 307
DB 165 -----SEVTE-VKI-EVKTEDSPPE-ESIT 187
QY 308 TPFSQCFCKESFPGPPIPLHQHERYLCKMNEIKAVLQPHENIVPNKAGYFVDNKA----- 362
DB 188 PAVSMGL-----SPAP-----EQNNEBMNNGSGSDCKSPDW 222
QY 363 -LLLSVSEKGLTSPINPYKDMSVLKAYVAMNMPNSDELLKISIVGLPQEPFKEMF 421
DB 223 PLRSRSLND-----SQVAVLQNHFKENPPFSKTELSAVALQIGVNRVYQVWF 271
QY 422 EQRKVQYSNRSPSLERTSKPLAPNSNPTTKDILLPRSPVKMDSITSPSIAELHNSVT 481
DB 272 QNTRAKERSNRPLPSIPGVSASAAA-----AATSTTWQ----- 307
QY 482 SCDPPLRLTK--SSHFTNIKAVDKLHRSRNTSPPLNLSSTSSKNSHSSSYTPNSPSSSE 539
DB 308 -TPVQLMAAMASQFSN-----GNNSLTASQDERNMENTDEVMHDHC 347
QY 540 L---QAEPLDLSLPKQMRPK-----GIATKTKKARS 570
DB 348 LKDGKTPDLTSTDTDEPWSPEKLIQGLDQTVGLQELRLQAGNGFVTDDEDEK 407
QY 571 INLDHNSVSSSEN-----SDEPLNLTFKKFSNNLND----- 606
DB 408 IKAEBSPVSGSSSIMPSTIGQYPSILDSASLSVLEKALDVST--DFSQKTFPDSIGR 465
QY 607 -----KSNMPVFGNPPSAKPLYPPLPQSAFPATMPVQTSIPGLRPPGL 655
DB 466 EDIRAVRYVLIYSNKQIF-----EKLYYP----- 489
QY 656 DQMSFLPMAYTPTGAATPADMOQRKRYKRGQFGODLLDGAQDYMSGLDMDTDSCL 715
DB 490 -----PRYALT-----NNFQDMQK-----SSBDA--SSLCS 515
QY 716 SRKKI-----KKTESGYACDLCDKTFQKSSLLRHKYEHTGKPHOCQICKKAFK 768
DB 516 NESKLLKFTPTPLKEEGLFSCQCDKVFQKSSLLRHKYEHTGKPHOCQICKKAFK 575
QY 769 HHLIHSRLHSGEKPYQCDKGRKFSHSGYSQHMNRYSYCKRAE 816
DB 576 HHLTEHRLHSGEKPYQCDKGRKFSHSGYSQHMNRYSYCKRPREQ 623

```

RESULT 10

```

Q8MSQ8 PRELIMINARY; PRT; 569 AA.
AC Q8MSQ8;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE SD06902P.
GN ZFH1 OR CG1322.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota, Metazoa, Arthropoda, Hexapoda, Insecta, Pterygota;
OC Neoptera, Endopterygota, Diptera, Brachycera, Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Fatfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,

```

```

RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragae V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY118654; AAM50023.1; -
DR FlyBase; FBgn004606; zfh1.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00046; homeobox.1.
DR Pfam; PF00096; zf-C2H2; 1.
DR Prodom; PD000010; Homeobox; 1.
DR Prodom; PD000003; Znf_C2H2; 1.
DR SMART; SM00389; HOX; 1.
DR SMART; SM00355; Znf_C2H2; 4.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 569 AA; 63968 MW; DC7141EPF2618B1 CRC64;

```

Query Match 10.7%; Score 533.5; DB 5; Length 569;
 Best Local Similarity 27.9%; Pred. No. 68-25;
 Matches 172; Conservative 68; Mismatches 187; Indels 189; Gaps 23;

```

QY 250 ATKSIIDYLEKVNKAACLOSLTDSRQISNIKEKRTLIDLVTDKMIENHS---I 306
DB 85 ATEAATPIKREBSREX-----SPDESYSRSSQAIKQEDPLNVAERQTPVEBHPVSH 139
QY 307 TPFSQCFCKESFPGPPIPLHQHERYLCKMNEIKAVLQPHENIVPNKAGYF----- 358
DB 140 SADLRCSRSCKFNPHTLVQHEKVLGCL--IKELEQH--FOOQQAISFALSASEED 194
QY 359 --DNKALLSVLSEKQ-----LTSPIPKDMVSLKAYVAMNMPNSDELLKISIVAG 411
DB 195 EEDDEMDVBEERQBSGKRVKRVRAIN--EQOQQLKHYSILNRPSSDERMTAAARQ 252
QY 412 LQEPYKEMF-----QKRVQYSNRS-----PLERTSKPLAENSPPTKD 454
DB 253 LDPVVQVQVFNRRSERRKQGFQNNQAAGAAPMPIDSOASLTRDDPL-----D 303
QY 455 SLLRSPVYKPMQISIPSIKAEHLNSVTSCDPLRLTKSHFTNIRKAVD-KLDHSRNTPS 513
DB 304 LSVKRPPLTPKSESSPPYIA-----PP-----SGEALNPEALINSRKFSTASWS 348
QY 514 PNLSTSSKNHSSSYTPNSPSSSELAQP-----LDLSLPKO-----MRBP 556
DB 349 PASISSSAALYFGAAPPSPPNSQLDSTPPSGAFLPGLPPYMLPMSLPMELPFMR-P 407
QY 557 KGIATKNTKATISNL-DHNSVSSSSENSEDEPLNLTFKKFSNNSNLDKSNNPVFGM 615
DB 408 GGDFAH-NHALNMSIKLPYRGTSLSPGGSE-----KRSWRDD----- 445
QY 616 NPFSAPLYTPLPQSAFPATMPVQTSIGLRYPGLDQMSFLPMAYTPTGAATF 675
DB 446 -----SRISHDEFGAVLMP----- 462
QY 676 ADMQRRKTYQKRGQFGODLLDGAQDYMSGLDMDTDSCLSRKKIKTSGMYACDLCDK 735
DB 463 ---KPRGALETHGAGD-----PDLP-----YVDCQCK 489
QY 736 TFOKSSLLRHKYEHTGKPHOCQICKKAFKXHLIHSRLHSGEKPYQCDKGRFSH 795
DB 490 AFAKQSLARHKKYEHTGKPHOCIECPKAFKXHLIHSRLHSGEKPYQCDKGRFSH 549
QY 796 SGGYSQHMNRYSYCK 811
DB 550 SGGYSQHMNRYSYCK 565

```

RESULT 11

```

Q8C5P2 PRELIMINARY; PRT; 189 AA.
ID Q8C5P2

```

AC Q8C8P2;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Zinc finger homeobox 1a.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs."
RL Mature 420:563-573(2002).
DR EMBL; AK077901; BAC37053.1; -
SQ SEQUENCE 189 AA; 21400 MW; 52DEB7A580E9E978 CRC64;

Query Match 7.3%; Score 365; DB 11; Length 189;
Best Local Similarity 44.6%; Pred. No. 3.8e-15;
Matches 82; Conservative 27; Mismatches 47; Indels 28; Gaps 6;

QY 776 RLHSGEKYQCDKCKRFSHSGSYSGQNMNHRYSYCKREAREAREAREKGLGPTL 835
DB 2 RLHSGEKYQCDKCKRFSHSGSYSGQNMNHRYSYCKREAREAREAREKGLGPTL 56
QY 836 --LNNRAYLQSTPGYSDSEERESMP--DGESEKHEKGE--GYGK-LRRDGD 886
DB 57 PEVLATVHVGARASPSQADSDERESLTREBEDSEKEEEDKEMELQEGKECENPOGE 116
QY 887 EEEEEEESEENKSDTDPETIRDEEFTGDH-----SMDDSSDGMENKKS 932
DB 117 EEEEEEESEEEEEEVEADEAHEAAAKTGTVEVGAQAQAGSLQKASESESESES 176

QY 933 DHEE 936
DB 177 ESEQ 180

RESULT 12
Q8C827 PRELIMINARY; PRT; 914 AA.
AC Q8C827;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Zinc finger protein 62.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs."
RL Mature 420:563-573(2002).
DR EMBL; AK048619; BAC33395.1; -
SQ SEQUENCE 914 AA; 104811 MW; 196A50D47205E9B9 CRC64;

Query Match 6.9%; Score 343.5; DB 11; Length 914;
Best Local Similarity 21.5%; Pred. No. 6.7e-13;
Matches 196; Conservative 125; Mismatches 324; Indels 265; Gaps 41;

QY 6 AGNRKFKTECGAKAFKRYHKLHLRLHSGEKYECPCNCKKRFSSGYSYSHISKKGIC 65
DB 210 SGEKNCCKDECGKSFNYSVLDQHKRIHIGKEKPYECGCGKAFRNS-----G 257

QY 66 LISVNGRRNNIKTGSSP---NSVSSSEPTNSAITOLRNKLENGKPLMSBQTG----- 115
DB 258 L-----RVNHRKIHTGKEKPYECDDTGKTFSSNSGRLVHNRKIHTGKEKPYECDCGKAFITCR 312
QY 116 -LTKITEPLDND--KYVLMATGFGSSGPPFNNGLGATSPLVGHPAGSPMOLGYGM 172
DB 313 TLNHRKS--IHFGKPKYCDCEKSFNYSILLI-----QHKVHTG--- 351
QY 173 EAPLLGPEPTMNSLSEVOKVLO-----IVDVTNRQKMDCKTEDISKLGVMKDPGSP 227
DB 352 EKFY-----ECDECGKAFRNSGGLVHNRKIHTGKEKPYCDICGKAFSTSS----- 396
QY 228 EEOGVTSPNIPVGLPVVSHNGATKSIIDYTLKVNKAQLOSLTDSRQISNIKKEK 287
DB 397 -----GLAV-----HKSI--HPGKAHEKDCGKSPNYS----- 424
QY 288 LRTILDVTDKMLBNHSISTPFSCQPKESFPQGPPIPLHGERLYCKNBEIKAVLPHE 347
DB 425 -----LILQHKTI--HTGERPYVCDVCGKTFRNSGKLVHRL-----HT 462
QY 348 NIVPNKAGVFVNDKALL--LSSVLSKGLTSPINPK-----DHMSVLKAYYAMNMP 398
DB 463 GEKPKCDVC--GKAYISRSLKNHKGIMGEKPKYSYCKSFNYSALBQHKRIHTR 520
QY 399 ---NSDELKISIAVGLPQEFVKEMFEQR--KYVQ--YSNRSRPSLERTSKPLAPNSP 450
DB 521 KPFQDCBGKA-----FRNSGLKXHKRIHGERPYKCECGKAYISLSL 566
QY 451 TTQSLDLP-RSPVKPMS-----ITSPSIAELANSVSCDPLR--LTKSHFTNIKAVD 502
DB 567 INHKSVMHGEKPFKDCCEKAFITRYTLN--HKIHLQEKPKYCDVCKSFNYSLSLQ 625
QY 503 KLDSRSSTPRLNLSISSKNSHSSYTPNPSFSEBELQAPRLDLSLPKQREPRGIAT 562
DB 626 KRVHTR--KPFECRCVKYFRNNS-----LKVHKI----- 656
QY 563 KNTKATISINDHNSVSSSSSENSDEPLNLTFIKEF--SNSNLDNKSNNPFGMNPFAK 621
DB 657 -----HTGEKPYECDDICGAYISHSLINHKSTHP-----CK 688
QY 622 PLYTLPPOSAPPATMPPVQTSIPGLRPYGLDQMSFLPHMATYPTGATFADMQOR 681
DB 689 TSYTCDECGKAFSSRTLISHKRVHLGKPPKVE-----CGKSPFSYSLLS---QHK 738
QY 682 RKQKQGFQGLDLDAQDYMVGGLDMDTSDSCLSRKIKITKTESMAYCDLCTFQSS 741
DB 739 RIHTGKEKPYVCDMGKAFRNSGL-----TVNKRKIHTGK--PYGDCCEKAYISHS 788
QY 742 SLNRKYEHTGKRPPOCOICKKAFKHKHLLESHRLHSGEKPYQCDKCKRFSHSGYSQ 801
DB 789 SLINHKSVHRGKQPNCE--CGKSPNYSVLDQHKRIHIGKPKYRNDGCGKAFNIRSNLTk 847
QY 802 HNNHRYSYCKREAREAREAREKGLGPTELMNNRAYLQSTPGYSDSEERESMPR 861
DB 848 H-----KRIHTGEESLNNAMNESHSGTFQKMI---YEG-----GNSLDGTRMQNP- 890
QY 862 DGESEKHEK 871
DB 891 --VWEAEHAK 898

RESULT 13
Q35700 PRELIMINARY; PRT; 949 AA.
AC Q35700;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Eviddelta 105.
OS Evi1 OR Evi-1.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCB1_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bartholomew C., Clark A.M., Kilbey A.:
 RT "Identification and characterization of novel murine Evi-1 proteins.";
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=8311086; Pubmed=284206;
 RA Morishita K., Parker D.S., Mucenекi M.L., Jenkins N.A., Copeland N.G.,
 RA "Retroviral activation of a novel gene encoding a zinc finger protein
 RT in IL-3-dependent myeloid leukemia cell lines."
 RT Cell 54:831-840(1998).
 DR EMBL; AJ001482; CA04777.1; -
 DR HSP; P08047; 1SP2.
 DR MGD; MGI:95457; Evi1.
 DR InterPro; IPR003006; Iq_MHC.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00096; zf-C2H2; 10.
 DR Prodom; PD000003; Znf_C2H2; 4.
 DR SMART; SM00355; Znf_C2H2; 10.
 DR PROSITE; PS00280; Iq_MHC; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 10.
 DR Metal-binding; zinc; zinc-finger.
 KW SEQUENCE 949 AA; 106536 MW; 50754E1A26A158BB CRC64;

Query Match 6.8%; Score 341.5; DB 11; Length 949;
 Best Local Similarity 19.8%; Pred. No. 9,4e-13;
 Matches 211; Conservative 133; Mismatches 354; Indels 369; Gaps 44;

OY 9 RKFKTECGKAFYKHKHLKEHLRHSGEKYECPCNCKRPSHSGSYSHISSKKGICGLIS 68
 DB 101 RKFKTECGKAFYKHKHLKEHLRHSGEKYECPCNCKRPSHSGSYSHISSKKGICGLIS 155
 OY 69 VNGRMNNIKTGGSPNSVSSPTNSAITQLRNKLKENGKPL-----SMSEGTGLIKIT 121
 DB 156 -----GAAHACPECCKTFATSSGKQKHHSVKPFICEVCHKSYTQFNSLCKHKR 208
 OY 122 EPLD-----PNDYKVLMTATGFSG-----SSPPM----- 145
 DB 209 MHADCRTOIKDCQGMFTSTSLINKRRCEGKNHFAAGFFGQGISLPTGTPAMDKTSM 268
 OY 146 -----NGGE-----GATSP-----GVHPSAQSPMQHLGVMEAPLIGF 179
 DB 269 VNNSHANPGLADYFGTRHPAGLTFPTAPGFSFSGFLPFS-----GLYHRPPLIPA 320
 OY 180 PTNNSLSEVQKVLQIVDNTVSRQKMDCKTEDISKLGYHMKDPCSQPEQGVTSPIPI 239
 DB 321 SPVKGSLSTEGSNKCGSPLLTQHQLPATQDILKALSKI-----PP 362
 OY 240 VG-----LPVYSHGATSIIDYTEKNE-AKACLOSLTDSRRQISNKEKLT 230
 DB 363 VGDNKPVLELP-----ERSSSERPLEKSIDSESSDLDVSTPSSGSDLETSSGDLIES 415
 OY 291 LILVLTDDMKIEMHSISTPSCQCFSPGPPIPLQHHERYLCQMEIKAVLOPHENI- 349
 DB 416 --DIESDKEK-----CKEN-----GKMFYDKTSPLQNTLA 442
 OY 350 -VENKAGVFNKALLLSVLSSEKGLTSPIN-PYKDHMSYLKAYYANMPEPNSDELKIS 407
 DB 443 SINNKKE--HNNSVFSASVEGSAVGSIDKAIASIAEKYF-----GSGVGLGLO 494
 OY 408 I-----AVGLDQEFYKWF--QRKVQYNSRSRSPSLERTSKPLAPNSNPTTKOSLPRSP 461
 DB 495 DKRYGALPYPSMPLPFPPAFPSQSMYPF-----PDRDLRSIPL-----KMEQSP 539
 OY 462 VKRMDSTSPSIAMHNSVTSQCPPLALTKSSHFTNIKAVDKLDHNSNTSPPLNLSSTS 521
 DB 540 -----SEVKKLQKSSSESPPDLITRKD 562

OY 522 SKNSHSSSTYSPNSFSSSELOAPPLDLSLPKQREPKGIATQNKTKATSIINDHNSVSS 581
 DB 563 EKPITSGPSKPSGTPATS-QDQPLDSMGSRG-----ASGTYLTPRRKHHFGEKK 613
 OY 582 SENSND-EPLNLTFFIKKEFENSNNLDNKSNNPFGMNP-----SAKPLTYLPPQSAVPAT 637
 DB 614 GSNMIDRP-----SSDGLQHARETFPF-MDPIYRERKLTLDLELKE----- 657
 OY 638 FMPVQTSIPGLRPYPGLDQMSFLPHMAATYPTGATFADMOQRKYORKQGFQGLDGL 697
 DB 658 -----KYLARPSG-----FLTHPQPLP-----DQR----- 678
 OY 698 AADYNSGLDDMTDSDSCSRKIKKTE-----SCMYAD 731
 DB 679 --TWMSAINMAEKESFALPPEASSELLQSVSPMSFPAPNTLPENLRKGRKERTYCR 736
 OY 732 LCDKTFQKSSSLRHKEHTGKRPQCOICAKAFKHLHBSR-LHSGEKPYQCDKCG 790
 DB 737 YCGKTFPRANLTRHRTHTGEOYRCKYCDRSFSSISMLQHVNNIHNKEKPKCHLCD 796
 OY 791 KPFSSGYSOHHNHRYSYCKREAREAREAREKGH/GPTLLMNRAYLQSIITPGY 850
 DB 797 RCGGQGTNLDRHL-----KHHNGMSGT-----ATSSPH 826
 OY 851 SOSERESEMPRQSEKEHEKE--GEQYGLRRRQDGESESESESESESESESESESESESE 905
 DB 827 SELESAGAILDDKEDAYFTEIRNFIGNSNHGSQSPRN-MEEMNGSHPFQKVALTSQNS 885
 OY 906 ETIRDEEFTGDHSDMDSSED-----GKME-----TSDHE--EDNMED 941
 DB 886 DLUDDEVEDEVALDEDEEDNDIPGKPKRELGVTRLDEIPEODYE 932

RESULT 14

OQRS51

ID OQRS51 PRELIMINARY; PRT; 743 AA.

AC OQRS51;

DT 01-JUN-2002 (Tremblrel. 21, Created)

DT 01-JUN-2002 (Tremblrel. 21, last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, last annotation update)

DE Hypothetical 84.9 kDa protein.

GN ZFP62.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

RN NCB1_Taxid=10090;

RP SEQUENCE FROM N.A.

RA Strausberg R.;

RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC022935; AA822935.1; -

DR MGD; MGI:99662; Zfp62.

DR InterPro; IPR007087; Znf_C2H2.

DR Pfam; PF00096; zf-C2H2; 22.

DR SMART; SM00355; Znf_C2H2; 22.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 22.

DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 22.

KW Hypothetical protein; Metal-binding; zinc; zinc-finger.

KW SEQUENCE 743 AA; 84942 MW; D1482B5602E5216 CRC64;

Query Match 6.8%; Score 331.5; DB 11; Length 743;
 Best Local Similarity 20.3%; Pred. No. 2,8e-12;
 Matches 166; Conservative 89; Mismatches 254; Indels 307; Gaps 29;

OY 6 AGRKFKTECGKAFYKHKHLKEHLRHSGEKYECPCNCKRPSHSGSYSHISSKKGICG 65
 DB 203 SGKKNCKCEGSGSFYSSVLDOHKRIHTEKPECEGCKARNS-----G 250
 OY 66 LISVNGRMNNIKTGGSP---NSVSSPTNSAITQLRNKLKENGKPLMSBQTS----- 115
 DB 251 L-----RVKRIHTGKAPYECOTCGTFPSNSGLRHKHGIHTEKPECEGCKARITCR 305
 OY 116 -ILKITEPLDFND--YKVLMTATGFSGSSPFNNGLAGATSPGLVHPSAQSPMQLGVGM 172

```

Db      TLNHS--HFQDPRYKDECEKSFNYSLLI-----QHKVHTG-- 344
Qy      173 EAPLIGFPTMSNLSVQKVLQ-----IYDNTVSROKMDCKTEDISKLGYNMKDPCSGP 227
Db      345 EEPY-----ECBECGKAFRNSGLYHKRIHGEKRYKCDICGKAFSYS----- 389
Qy      228 EEOGVSPNIPVGLPVVSHNGATYSIIDYLEKVNKAACLOSLTDSRROIISIKKEK 287
Db      330 -----GLAV-----HKSI--HPGKAHECDCKSFYSN----- 417
Qy      268 LRTLIDLVTDKMIENHSISTPSCQFCESFPGLPHQHERYLCNMKEIKAVLQPH 347
Db      418 -----LLQHKTI--HTGERPYCDVCGCKTFRNSGLKVHRL-----HT 455
Qy      348 NIVPKAGVFDVKNKLL--LSVISEKGLTSPINPYKDHSVLKAYANMEPNDELKI 406
Db      456 GEKPYKCYVC--GKAYISHSLSKHKIGIMGEKPYKCYCEKSFYSALB----- 504
Qy      407 SIAGVLPQEFYKEMFEQRKVQYNSRSPSLERTSKPLAPNSNPTTKOSILPRSPVKPM 466
Db      505 -----QHKRI--HTREKPRGDCGCKAFRNSG----- 530
Qy      467 SITSPSIABLNSVTSQDPLRLTKSSHFTNIKAVDKLDHSRNTPSPLNSTSSKNSH 526
Db      531 -----LKVHRIHTGERPYKCECG--KAY-----ISLSLHNKSV 565
Qy      527 SSSYTPNSFSSEELQAEPLDLSLPKOMREPKGIIATKNTKATSNLDHNSVSSSENSD 586
Db      566 HGEKAP--FKDCE-----KAFITRYTLNKKIHL-----GE 597
Qy      587 EPLNLTFLIKKEPSNNSNLDKNSNPVFGMNPFSAPLPTPLPQSAFPPTATMPVQTSI 646
Db      598 KPYKCDVCEKSFN-----YTSLL----- 615
Qy      647 PGLRPYGLDQMSFLPHMAVYTPGATPADMQQRKTKQKQFGQDLIDGADYMSGLD 706
Db      616 -----SQRKVTREKPFED-----RCE 634
Qy      707 DMTDSDCLSRKKIKKTESGMVACDLCDKTFQKSSSLRHKYEHTGKRPQOQICKKAF 766
Db      635 KVFRRNSSLKHKRIHTEKPYECDICGAYISHSLSLHNKSTHPGKSYTDCGKAF 694
Qy      767 HKHHLIEHSRLHSGEKPYQCDKGRFSGSGYSQH 802
Db      695 SSRTLISHKRVHLGKPFKVCBGKSFYSLSLQH 730

RESULT 15
Q8TB9 PRELIMINARY; PRT; 751 AA.
AC Q8TB9;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC022992; AAH22992.1; -.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; zf-C2H2; 19.
DR ProDom; PD000003; Znf_C2H2; 18.
DR SMART; SM00349; KRAB_1.
DR SMART; SM00355; Znf_C2H2; 19.

```

```

DR PROSITE; PS50805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 19.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 19.
DR Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 751 AA; 86190 MW; A05FCECBF1275CB CRC64;

Query Match
6.6%; Score 331.5; DB 4; Length 751;
Best Local Similarity 18.1%; Pred. No. 2.9e-12;
Matches 148; Conservative 77; Mismatches 198; Indels 393; Gaps 27;

Qy      7 GNRKFKTECGKAFYKHKHLEHLIHSGEKPYECPNKKRPSHSGYSYSHSKKICGL 66
Db      302 GEKPYKCECGKAFSQRTHLVQHORIHTEKPYTCNECGKAFSQRGHMEH----- 352
Qy      67 ISVNGRMNNTIKTSSP--NSVSSPNSAITOLRNLTKNGKPLMSBQGLTKITP 123
Db      353 -----QKIHTEKPYKCECDKTFTRSHLTQHKIHGE-----KT-- 389
Qy      124 LDFNDKYVLMATHGFSGSSPMMNGIGAT--SPLGVHPSAOSPMQHLGVMEAPLLGPPT 181
Db      390 -----YKNECGKAFNPGSTFIRHMIHTEKPYECNECGKAFSQH----- 430
Qy      182 MNSNLSVQKVLQIYDNTVSRQKMDCKTEDISKLGK-----HKMDPQSQPEEGVTS 235
Db      431 --SNLTHQK-----THTGEKPYDC--ABCGKSPYSWSIAQHLK----- 466
Qy      236 NIPVGLPVSHNGATKSIIDYLEKVNKAACLOSLTDSRROIISIKKEKLTFLIDL 295
Db      467 -----IHTGEK--PYKNECGKAFSYCSLT--QHKRI----- 495
Qy      236 TDDKMIENHSISTPSCQFCESFPGLPLHQRH-----YLCNMKEIKAVLQPHEN 348
Db      496 -----HTREKPREGCECGKAFSYLSNLNQHKHTHQEKAYEBK----- 533
Qy      349 IVPKAGVFDVKNKALLSSVSE--KGLTSPINPYKDHSVLKAYANMEPNDELKIS 407
Db      534 --ECG-----KAFIRSSSLAKHERIHTGEKPYQCH----- 561
Qy      408 IAVGLPQEFYKEMFEQRKVQYNSRSPSLERTSKPLAPNSNPTTKOSILPRSPVKPMD 467
Db      562 -----EGKTFSTGS-----SLIQ----- 575
Qy      468 ITSPSIABLNSVTSQDPLRLTKSSHFTNIKAVDKLDHSRNT--PSPLNSTSSKNSH 526
Db      576 -----HKRIHTEKPYKNECGKAFN--QNHLTQHKRIHTEKPYECGKAFRH 625
Qy      527 SSSYTPNSFSSEELQAEPLDLSLPKOMREPKGIIATKNTKATSNLDHNSVSSSENSD 586
Db      626 CSS-----LAHQHT-----HTE 639
Qy      587 EPLNLTFLIKKEPSNNSNLDKNSNPVFGMNPFSAPLPTPLPQSAFPPTATMPVQTSI 646
Db      640 KPYQCNKCEKTFSSQSHLTQHOR-----IHT----- 665
Qy      647 PGLRPYGLDQMSFLPHMAVYTPGATPADMQQRKTKQKQFGQDLIDGADYMSGLD 706
Db      666 -GEK----- 669
Qy      707 DMTDSDCLSRKKIKKTESGMVACDLCDKTFQKSSSLRHKYEHTGKRPQOQICKKAF 766
Db      670 -----YKNECGKAFSQRTHLTQHORHTGKPYKNECGKTF 708
Qy      767 HKHHLIEHSRLHSGEKPYQCDKGRFSGSGYSQH 802
Db      709 QSTYLIQORLHSGEKPGCNDGKSFYRSALNKH 744

```

Search completed: January 5, 2004, 16:31:02
Job time : 54 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: January 5, 2004, 16:30:01 ; Search time 21 Seconds
(without alignments)
1901.974 Million cell updates/sec

Title: US-09-964-238-2

Perfect score: 4991
Sequence: 1 MLTQAGNRKFKTECGKAF.....DGKMETKSDHEEDNEDMG 944

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA*
1: /cgn2_6/ptodata/1/aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4991	100.0	944	US-09-449-285A-2	Sequence 2, Appli
2	276.5	5.5	706	US-08-074-967-2	Sequence 2, Appli
3	276.5	5.5	706	US-08-553-541B-2	Sequence 2, Appli
4	276.5	5.5	706	US-09-268-202-2	Sequence 2, Appli
5	276.5	5.5	706	PCT-US94-0669-2	Sequence 2, Appli
6	274	5.5	496	US-08-224-482-2	Sequence 2, Appli
7	274	5.5	533	US-08-040-548-1	Sequence 1, Appli
8	274	5.5	533	US-08-466-344-1	Sequence 1, Appli
9	274	5.5	533	5206152-2	Patent No. 5206152
10	272.5	5.5	543	US-08-224-482-4	Sequence 4, Appli
11	272.5	5.5	543	US-09-702-705-337	Sequence 337, App
12	272.5	5.5	543	US-09-736-457-337	Sequence 337, App
13	258	5.2	51	US-09-449-285A-21	Sequence 21, Appli
14	256.5	5.1	1706	US-08-459-568-2	Sequence 2, Appli
15	256.5	5.1	1706	US-08-399-411-2	Sequence 2, Appli
16	256.5	5.1	1706	US-08-516-859A-2	Sequence 2, Appli
17	256.5	5.1	1706	US-09-586-472-2	Sequence 2, Appli
18	256.5	5.1	1706	US-09-528-706-2	Sequence 2, Appli
19	248.5	5.0	1719	US-08-459-568-4	Sequence 4, Appli
20	248.5	5.0	1719	US-08-399-411-4	Sequence 4, Appli
21	248.5	5.0	1719	US-08-516-859A-4	Sequence 4, Appli
22	248.5	5.0	1719	US-09-586-472-4	Sequence 4, Appli
23	248.5	5.0	1719	US-09-528-706-4	Sequence 4, Appli
24	248	5.0	675	US-08-317-522A-9	Sequence 9, Appli
25	248	5.0	675	US-08-439-818A-9	Sequence 9, Appli
26	248	5.0	675	US-08-751-965-9	Sequence 9, Appli
27	248	5.0	675	US-08-738-975-9	Sequence 9, Appli

28	248	5.0	675	2	US-08-728-626-9	Sequence 9, Appli
29	248	5.0	675	3	US-08-808-599A-9	Sequence 9, Appli
30	235.5	4.7	387	2	US-08-224-482-8	Sequence 8, Appli
31	234.5	4.7	671	3	US-09-121-321-16	Sequence 16, Appli
32	234.5	4.7	671	3	US-08-933-803A-16	Sequence 16, Appli
33	230.5	4.6	488	2	US-08-933-750C-17	Sequence 17, Appli
34	230.5	4.6	488	3	US-09-234-613-17	Sequence 17, Appli
35	228.5	4.6	543	4	US-09-362-123A-4	Sequence 4, Appli
36	227.5	4.6	455	2	US-08-224-482-6	Sequence 6, Appli
37	227.5	4.6	456	1	US-08-040-548-2	Sequence 2, Appli
38	227.5	4.6	456	1	US-08-466-344-2	Sequence 2, Appli
39	226	4.5	728	2	US-08-475-844-5	Sequence 5, Appli
40	226	4.5	728	5	PCT-US95-08429-5	Sequence 5, Appli
41	221.5	4.4	727	2	US-08-475-844-9	Sequence 9, Appli
42	221.5	4.4	727	5	PCT-US95-08429-9	Sequence 9, Appli
43	220	4.4	470	2	US-08-946-241B-2	Sequence 2, Appli
44	220	4.4	470	3	US-09-309-053-2	Sequence 2, Appli
45	220	4.4	479	2	US-08-946-241B-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-09-449-285A-2
; Sequence 2, Application US/09449285A
; Patent No. 6313280
; GENERAL INFORMATION:
; APPLICANT: Vlaams Interuniversitair Instituut voor Biotechnologie
; TITLE OF INVENTION: SWAD-INTERACTING POLYPEPTIDES AND THEIR USE
; FILE REFERENCE: 2676-42320S
; CURRENT APPLICATION NUMBER: US/09/449,285A
; CURRENT FILING DATE: 1999-11-24
; PRIOR APPLICATION NUMBER: PCT/EP98/03193
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 97201645.5
; PRIOR FILING DATE: 1997-06-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 944
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-449-285A-2

Query Match	100.0%; Score 4991; DB 4; Length 944;
Best Local Similarity	100.0%; Pred. No. 0;
Matches	944; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MLTQAGNRKFKTECGKAFKJHLKHLRIHSGEKPECPNCKKRFSSHSYSSHS 60
DB	1 MLTQAGNRKFKTECGKAFKJHLKHLRIHSGEKPECPNCKKRFSSHSYSSHS 60
QY	61 KKGIGLISVNGRMNNITGSSPNSVSSPTNSAITQLRNKLKNGKPLSMSEDTGLKIK 120
DB	61 KKGIGLISVNGRMNNITGSSPNSVSSPTNSAITQLRNKLKNGKPLSMSEDTGLKIK 120
QY	121 TEPLDENYKYLMAHFGSSGSPFNNGLGATSPGVPSASPMOHGVCNEAPLLGFP 180
DB	121 TEPLDENYKYLMAHFGSSGSPFNNGLGATSPGVPSASPMOHGVCNEAPLLGFP 180
QY	181 TMNSLSEVQKVLQVDTVTRQKMDCTEDISKLKGYHMDPCSQPEOGVTSNIPV 240
DB	181 TMNSLSEVQKVLQVDTVTRQKMDCTEDISKLKGYHMDPCSQPEOGVTSNIPV 240
QY	241 GLPVVSHNGATKSIIDYTLKRVNEAKCLQSLTDSRRQISNIKKELRTLLDVTDDKM 300
DB	241 GLPVVSHNGATKSIIDYTLKRVNEAKCLQSLTDSRRQISNIKKELRTLLDVTDDKM 300
QY	301 IENHSISPFSCQFKESFPGPPIPLQHHERYLCNNEIKAVLOQHEHIVPKAGVFDN 360
DB	301 IENHSISPFSCQFKESFPGPPIPLQHHERYLCNNEIKAVLOQHEHIVPKAGVFDN 360

Qy	361	KALLLSVLSBGLSPINPYQDNHSLVLAAYAMNPNPSDELLKISIAVGLQOEYKEM	420
Db	361	KALLLSVLSBGLSPINPYQDNHSLVLAAYAMNPNPSDELLKISIAVGLQOEYKEM	420
Qy	421	FEQRKVQYSNRSPSLERTSKPLAPNSNPTTKDLSLPPSPVYKPMDSITPSSIAELHNSV	480
Db	421	FEQRKVQYSNRSPSLERTSKPLAPNSNPTTKDLSLPPSPVYKPMDSITPSSIAELHNSV	480
Qy	481	TSCBPLPLTKSSHNTNIAVDKLDHSRNTB9PLMLSSSTSKNHSSTPNSPSSSEL	540
Db	481	TSCBPLPLTKSSHNTNIAVDKLDHSRNTB9PLMLSSSTSKNHSSTPNSPSSSEL	540
Qy	541	QAEPLDLSLPKQMRPEKGIITATKNKTKASINLDHNSVSSSSNSBEPNLATPKKEFSN	600
Db	541	QAEPLDLSLPKQMRPEKGIITATKNKTKASINLDHNSVSSSSNSBEPNLATPKKEFSN	600
Qy	601	SNNLDNKSNNPVFGNMPFSAKPLYPPLPOSAPFPATFMPVOTSIPLGLRYPGLDOMSF	660
Db	601	SNNLDNKSNNPVFGNMPFSAKPLYPPLPOSAPFPATFMPVOTSIPLGLRYPGLDOMSF	660
Qy	661	LPHMAVYTPGATPADMOQRKRYQKQGFQGLDGAODYNSGLDMDTSDSCLSRKTI	720
Db	661	LPHMAVYTPGATPADMOQRKRYQKQGFQGLDGAODYNSGLDMDTSDSCLSRKTI	720
Qy	721	KKTESGMAACDLCDTFPQKSSSLRHKYBHTGRRPHQCOICCKAFYHKHLLIHSRSHSG	780
Db	721	KKTESGMAACDLCDTFPQKSSSLRHKYBHTGRRPHQCOICCKAFYHKHLLIHSRSHSG	780
Qy	781	EKPYQCDKCGKRFSSHSGSYQGMNHNYSYCKEAEEREAAERAREKGHGLPTELLMTRA	840
Db	781	EKPYQCDKCGKRFSSHSGSYQGMNHNYSYCKEAEEREAAERAREKGHGLPTELLMTRA	840
Qy	841	YLQSTIPQGYSDSEERESMPRDGESHKEHKGEBEGYGLKLRDDGDEEBEESSENKS	900
Db	841	YLQSTIPQGYSDSEERESMPRDGESHKEHKEGESEGYGLKLRDDGDEEBEESSENKS	900
Qy	901	MDTDEETIRDEEETGDHSMDDSSDQCKMETKSHEDENNEDMG 944	
Db	901	MDTDEETIRDEEETGDHSMDDSSDQCKMETKSHEDENNEDMG 944	

RESULT 2
US-08-074-967-2
Sequence 2, Application US/08074967
Patent No. 561672
GENERAL INFORMATION:
APPLICANT: Dalla-Favera, Riccardo
APPLICANT: Chaganti, R.S.K.
TITLE OF INVENTION: CLONING AND USES OF THE GENETIC LOCUS
TITLE OF INVENTION: bcl-6
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/074,967
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/43771
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550

```

? TELEFAX: (212) 664-0525
? TELEF: 422523 COOP UI
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 706 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-074-967-2

```

Query Match	5.5%;	Score 276.5;	DB 1;	Length 706;
Best Local Similarity	21.1%;	Pred. No. 2,1e-13;		
Matches	126;	Conservative	87;	Mismatches 244; Indels 141; Gaps 22
Qy	259	LEKVNKAACLQSLTTSRRQISNIK--KEKRLTLIDLIVTD-----KMIENH-----	304	
Db	112	LQMEHVVDTCRKFKIASEAEWVSAIKPPREFEFLNSRMIMPODIMAYAGREVENNLLPLRS	171	
Qy	305	-----SISTPSCQFCESPFQPIPLMOH--ERLCKNNERIKAVLOPHE	347	
Db	172	APGCEBAFAPASLYSGISTP-----PASYMYSHLVSSLLPSDEEFRIIDRMFVA	221	
Qy	348	NIVENKAGVFDNKALLLSVLSSEKQ---LTPSINPYKDHMSYLKAYANMEBNSDEL	403	
Db	222	NPFKE-----RALPDQSAKRPVGGYSRRTLEVPVCHSNLYSPKETIPEARSQ--	272	
Qy	404	LKISIAVLG---PQEFVKEMFEQRKYQYQISNSGSPSLERTSKPLAPNSNTTQDGL-P	458	
Db	273	MHYVAAGELKPAASASANAPYFPQDKASK--EERPPSSDEDEIHALHEPPNAPLNRKGLVSP	331	
Qy	459	RSPPK-----PMDSLTSPSIAELHNS-----VTSQCPRLTLKSSHFNIKAVDKLDH	506	
Db	332	QSPKSDCQPNSPLEACSSKNACILQSGSPAKSPDPDKACSMKYYEYIUNSLNQ--N	389	
Qy	507	SRNSTPAPLNLSTSSKNSSHSSYTPNPSFSSEELQAEPLDLSLPKQMEEPKGIATKNT	566	
Db	390	AKPGPQDAELGRLSPPR-----AVTAPACQPMPEPVLQLQSPYKL-----	431	
Qy	567	KATSLINDHNSVSSSSSENSDEPLNLTLTKKEFNSNNLIDKSNNPVFGMNPFSAPKPLTYP	626	
Db	432	-----SASGDBSTIP-----QASRLNINVMNMTG---SPRSSSESHSP	467	
Qy	627	LPPQSAFPAPATFM--PVPQTSIPGLRPVGLDQMSFLPMAVTPYTGATFPADMOQRKXQ	685	
Db	468	L-----YMPPKCTSGGSSP-----QHAEMCLTAGETPAE---EKGE	503	
Qy	686	RKQGFQDILLDGAODYMSGLDMDTSDSCLSRKIKIKTESGMYACDLCDTKFOKSSLLR	745	
Db	504	TQSEYSDSSCENAGAFPCNECDSPSEBSASLKRHTLQHTSDPKRYCDRCQAFRYKGNLAS	563	
Qy	746	HKYHTGKRPQOCCICKKAFKRNHLLHNSLHSGEYKQOCDKSGKSPSHSGYSQHM	803	
Db	564	HKYVTEKEPRNCICGAQFNRPNMLKTHTRHSGEYKQKQETGAGRRVQVUANLRHV	621	

RESULT 3
 US-08-553-541B-2
 : Sequence 2, Application US/08553541B
 : Patent No. 5882858
 :
 : GENERAL INFORMATION:
 : APPLICANT: Dalla-Pavera, Riccardo
 : APPLICANT: Chaganli, Raju S.K.
 : TITLE OF INVENTION: CLONING AND USES OF THE GENETIC LOCUS
 : TITLE OF INVENTION: bcl-6
 : NUMBER OF SEQUENCES: 9
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Cooper & Dunham LLP
 : STREET: 1185 Avenue of the Americas
 : CITY: New York
 : STATE: New York
 : COUNTRY: United States of America
 : ZIP: 10036
 :
 : COMPUTER READABLE FORM:
 :

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,541B
FILING DATE: May 28, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/43771-A-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELETYPE: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-553-541B-2

Query Match 5.5%; Score 276.5; DB 2; Length 706;
Best Local Similarity 21.1%; Pred. No. 2.1e-13;
Matches 126; Conservative 87; Mismatches 244; Indels 141; Gaps 22;

259 LEKNEAKACLOSITTSRROISNK--KEKLRITLIDVTD-----KMIENH----- 304
112 LQMEHVVDTCRKFIKASAEWWSAIKPREEFELNSRMLMPDIMA YRGEVENNLPLRS 171
305 -----SISTPSCQCKESFPGPIPLHOH--ERYLCNNBEIKAVLOPHE 347
172 APGESRAFAPSLYSGLSTP-----PASYSMTSHLPVSSILFSDERFDRVMEVA 221
348 NIVNKAQVFDNKKALLSVLSBK-----LTSPINPKDMSVLYKAYANMEBNSDEL 403
222 NPFPRE-----RALPCDSARPYGYSRPTLEVSFVCHSNIVSPKETIPEEARSD-- 272
404 LKISIAVL-----POEFYKEMFEQRKYQYNSRSPSLERTSKPLAPNSPTTQDLSL-P 458
273 MHYVAELKRAAASARNAAPYPCDKASK--BEERPSSEDEIALHPEPPNAPLNKGLVSP 331
459 RSPVK-----PMDSTSPSIABLANS-----VTSCEPLRLTSSHFTNIKAVDKLDH 506
332 QSPKSDCQPNSTPEACSSKNACILQSGSPPAKSPTPKACSWKKYFIVLNSLNG--N 389
507 SRNTPSPPLNSTSSKNSHSSYTPNSFSSSELOAEPLDLSLPKQMEPKGIATKTKT 566
390 AKPGPEGAELGRLSR-----ATTAPACOPPMEPELIDQSPTKL----- 431
567 KATSLNDHNSVSSSENSDEPLNLTFTKKEFSNNLNDKSNPNVFGMNPSPAKPLTYP 626
432 -----SASGEDSTIP-----QASRLNINIVNSMTG-----SPRSSSESHSP 467
627 LPPOSAPPATFM--PPVOTSTIPGLRPPYGLDQMSFLPMATYTPYTGATPFADMOQRKYQ 685
468 L-----YMHPPKCTSCGSOSP-----OHAEMCLHTAGPTFAE-----EMGE 503
686 RKQGFQGLDLGAQDYMGLDMTSDSCLSRKKIKTESGMYACDLCDKTFQKSSSLR 745
504 TQSEYSDSSCENGAFCFNCBCDRFSEBASIKRHTLQTHSDKYKCDRQASFRYGNLAS 563
746 HKYHTTGKRPPOCQICCKAFKAKHLLIEHSRLHSEKRYQCDKGRPSHSGSYSGHM 803
564 HKTVTGKPYKCNICGAFNRPALKTHTRIHSGEKYKCTCGARFVQVAHLRAHV 621

RESULT 4
US-09-268-202-2
Sequence 2, Application US/09268202
Patent No. 6174997

GENERAL INFORMATION:
APPLICANT: Dalla-Favera, Riccardo
APPLICANT: Chaganti, Raju S.K.
TITLE OF INVENTION: CLONING AND USES OF THE GENETIC LOCUS
TITLE OF INVENTION: bcl-6
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/268,202
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/43771-A-PCT-US-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELETYPE: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-268-202-2

Query Match 5.5%; Score 276.5; DB 3; Length 706;
Best Local Similarity 21.1%; Pred. No. 2.1e-13;
Matches 126; Conservative 87; Mismatches 244; Indels 141; Gaps 22;

259 LEKNEAKACLOSITTSRROISNK--KEKLRITLIDVTD-----KMIENH----- 304
112 LQMEHVVDTCRKFIKASAEWWSAIKPREEFELNSRMLMPDIMA YRGEVENNLPLRS 171
305 -----SISTPSCQCKESFPGPIPLHOH--ERYLCNNBEIKAVLOPHE 347
172 APGESRAFAPSLYSGLSTP-----PASYSMTSHLPVSSILFSDERFDRVMEVA 221
348 NIVNKAQVFDNKKALLSVLSBK-----LTSPINPKDMSVLYKAYANMEBNSDEL 403
222 NPFPRE-----RALPCDSARPYGYSRPTLEVSFVCHSNIVSPKETIPEEARSD-- 272
404 LKISIAVL-----POEFYKEMFEQRKYQYNSRSPSLERTSKPLAPNSPTTQDLSL-P 458
273 MHYVAELKRAAASARNAAPYPCDKASK--BEERPSSEDEIALHPEPPNAPLNKGLVSP 331
459 RSPVK-----PMDSTSPSIABLANS-----VTSCEPLRLTSSHFTNIKAVDKLDH 506
332 QSPKSDCQPNSTPEACSSKNACILQSGSPPAKSPTPKACSWKKYFIVLNSLNG--N 389
507 SRNTPSPPLNSTSSKNSHSSYTPNSFSSSELOAEPLDLSLPKQMEPKGIATKTKT 566
390 AKPGPEGAELGRLSR-----ATTAPACOPPMEPELIDQSPTKL----- 431
567 KATSLNDHNSVSSSENSDEPLNLTFTKKEFSNNLNDKSNPNVFGMNPSPAKPLTYP 626
432 -----SASGEDSTIP-----QASRLNINIVNSMTG-----SPRSSSESHSP 467
627 LPPOSAPPATFM--PPVOTSTIPGLRPPYGLDQMSFLPMATYTPYTGATPFADMOQRKYQ 685
468 L-----YMHPPKCTSCGSOSP-----OHAEMCLHTAGPTFAE-----EMGE 503

QY 686 RKQFQGGDLLDGAQDYMSGLDMDTSDCLSRKKIKTESGMVACDLCDKTFQKSSSLR 745
DB 504 TQSEYSDSCENGAFPCNECDCCFSEASLKHHTLTQHSKDKPKCDRCQASFRYKGNLAS 563
QY 746 HKYHTGKRPHOCQICCKAFKHKHLLIEHSRLHSGKPYQCDKCGKRFHSGSYSQHM 803
DB 564 HKTVHTGKPYRCNICGAQFNRPNLKTHTRIHSGEKPYKCTCGARFQVAHLRAHV 621

RESULT 5
PCT-US94-06669-2
Sequence 2, Application PC/TUS9406669
GENERAL INFORMATION:
APPLICANT: Dalla-Favera, Riccardo
APPLICANT: Chaganti, R.S.K.
TITLE OF INVENTION: CLONING AND USES OF THE GENETIC LOCUS
TITLE OF INVENTION: bcl-6
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10112

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/06669
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/074,967
FILING DATE: 09-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 43771-A-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-06669-2

Query Match 5.5%; Score 276.5; DB 5; Length 706;
Best Local Similarity 21.1%; Pred. No. 2.1e-13;
Matches 126; Conservative 87; Mismatches 244; Indels 141; Gaps 22;

QY 259 LEKVNKAQALQSLTDSRQISNIK--KFKATLIDLVTD-----KMIENH----- 304
DB 112 LQMEHVVDTRKFKIKASEAMVSAIKPREEFINSRMLMPODIMAYVGRGVANNPLRS 171
QY 305 -----SISTPFSCOFCKESPFGPIPLHQH--ERYLCMNEIKAVLQPH 347
DB 172 APCGESARAFPSLYSGISTP-----PASYSMYSHLPVSSLLPSDEFPDVMFVA 221
QY 348 NIVPNKAGVFVDNKALLSSVLESG-----LTSPINDYKDMSVLAKAYYAMNEPNSDEL 403
DB 222 NPPPKR-----RALPCDSARVPVGEYSRPTLEVSPNVCHSNIVSPKETIPREARSD-- 272
QY 404 LKISIVNGL-----PQEVKEMFEORKYQYQNSNSRSPLETSKPLAPNSPTTKDILL-P 458
DB 273 MHYSVAEGLKPAAPASARNADYPFCDKASK--EEERPSSEDEIALHFEPNAPLNRKGLVSP 331

QY 459 RSPVK-----PMDSITSPSIAELHNS-----VTSCEPPLRLTKSSHFTNIKAVDKLDH 506
DB 332 QSPQSDCQPNPTEACSKNACILQSGSGPPAKSPPTDPAKCSWKYKXIVLNSINQ--N 389
QY 507 SRSNTPSPNLNSTSSKNSHSSSTYPNPSSELOLQAPLIDLSPLKQMEPKKILATKNT 566
DB 390 APGGEQALGLRLPR-----AYTAPACQPMPEPENLIDQSPYKL----- 431
QY 567 KATSLINDHNSVSSSENDEPLNLTFLIKKEFSNNSNLDKSNPNVPGMPPSAKLYTP 626
DB 432 -----SAGEBSTIP-----QASRLNNIVNRMTG-----SPRSSSESHP 467
QY 627 LPPOSAPPATFM--PVQTSIPGLRPPGLDQMSFLPHMAYTYTPGATFADMQRRXYQ 685
DB 468 L-----YMHPPKTSQSGSP-----QHMEMLHTAGPTFAE-----EMGE 503
QY 686 RKQFQGGDLLDGAQDYMSGLDMDTSDCLSRKKIKTESGMVACDLCDKTFQKSSSLR 745
DB 504 TQSEYSDSCENGAFPCNECDCCFSEASLKHHTLTQHSKDKPKCDRCQASFRYKGNLAS 563
QY 746 HKYHTGKRPHOCQICCKAFKHKHLLIEHSRLHSGKPYQCDKCGKRFHSGSYSQHM 803
DB 564 HKTVHTGKPYRCNICGAQFNRPNLKTHTRIHSGEKPYKCTCGARFQVAHLRAHV 621

RESULT 6
US-08-224-482-2
Sequence 2, Application US/08224482
Patent No. 5837692
GENERAL INFORMATION:
APPLICANT: Mercola, Dan
APPLICANT: Adamson, Eileen D.
TITLE OF INVENTION: Inhibition of the Mitogenic Activity of
TITLE OF INVENTION: PDGF by Mammalian EGR
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/224,482
FILING DATE: 07-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-ME 9913
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-224-482-2

Query Match 5.5%; Score 274; DB 2; Length 496;
Best Local Similarity 25.4%; Pred. No. 1.9e-13;
Matches 99; Conservative 47; Mismatches 143; Indels 100; Gaps 14;

QY 518 SSTGS-----KNSHSSYTPNSFSSER-----LQAEPL-DLSLPQMEPKKIITTKN 565

Db 27 SSTSGGGGGGSGSAGFNPQGEPSQPYEHLTTESFSDIALNNE---KAMVETSY 82
 QY 566 TKATISINDHNSVSSSSSENDEPL-----NLTFIKKEFNSNNLNDKSNPNVFGMNFSAK 621
 Db 83 SQTRL-----PPIYTGAFSLBEPANSGNTLMPBELFSLVGLVMTNPPSSSS----- 133
 QY 622 PLVPLPPOSAPFPATFMPVQTSIPG-----LRPYGLDQMSFLPHNAVITYPT 670
 Db 134 ---APSPASSSSSSASQSPPLSCAVPSNDSPISYAATFTPTMTD--IFEPQOQAPFG 188
 QY 671 GAATFADMQORRKYQKQGF-----OGDLIDGAODY--MSGLDMDTSDSC- 714
 Db 189 SAGTALQYPPPAYPATKGFQVPMIPDYLFPQOQGDLSLGTTPDQKPFQGLNRTQQPSLT 248
 QY 715 -LSRKIKKTESGM-----YAC--DLCDK 735
 Db 249 PLSTIKAFATQSGSDLKALNTTYSQSLIKPSRMKTPNRSKTPPHERPYACPVESCDR 308
 QY 736 TFOKSSSLRKHKEHTGKRPQCOICKAFKHKHLIHSRLHSGEKEPYQCDKCGKRFSH 795
 Db 309 RFSRSDLTRHIRHTGQKPFQCRICMNFSRSDHLTHIRHTGKEKPFACDICGRKFAR 368
 QY 796 SGYSQGMNHRYSYCKRAEBEREAEREA 824
 Db 369 SDEKRRHTKIHRLROKDKKXDVASPA 397

RESULT 7

US-08-040-548-1
 ; Sequence 1, Application US/08040548
 ; Patent No. 5763209
 ; GENERAL INFORMATION:
 ; APPLICANT: Sukhatme, Vikas P.
 ; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO THE
 ; NUMBER OF SEQUENCES: 67
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White & Durkee
 ; STREET: 321 No. 5763209th Clark Street, Suite 800
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: U.S.A.
 ; ZIP: 60610
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/040,548
 ; FILING DATE:
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Coughlin, Daniel F.
 ; REGISTRATION NUMBER: 36,111
 ; REFERENCE/DOCKET NUMBER: arc067
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (312) 744-0090
 ; TELEFAX: (312) 245-4961
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 533 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-08-040-548-1

Query Match 5.5%; Score 274; DB 1; Length 533;
 Best Local Similarity 25.4%; Pred. No. 2,1e-13;
 Matches 99; Conservative 47; Mismatches 143; Indels 100; Gaps 14;
 QY 518 SSTSG-----KNSHSSSYTPNPSFSSB---LQAEPL-DLSLPKQMEKPGIATKXK 565

Db 64 SSTSGGGGGGSGSAGFNPQGEPSQPYEHLTTESFSDIALNNE---KAMVETSY 119
 QY 566 TKATISINDHNSVSSSSSENDEPL-----NLTFIKKEFNSNNLNDKSNPNVFGMNFSAK 621
 Db 120 SQTRL-----PPIYTGAFSLBEPANSGNTLMPBELFSLVGLVMTNPPSSSS----- 170
 QY 622 PLVPLPPOSAPFPATFMPVQTSIPG-----LRPYGLDQMSFLPHNAVITYPT 670
 Db 171 ---APSPASSSSSSASQSPPLSCAVPSNDSPISYAATFTPTMTD--IFEPQOQAPFG 225
 QY 671 GAATFADMQORRKYQKQGF-----OGDLIDGAODY--MSGLDMDTSDSC- 714
 Db 226 SAGTALQYPPPAYPATKGFQVPMIPDYLFPQOQGDLSLGTTPDQKPFQGLNRTQQPSLT 285
 QY 715 -LSRKIKKTESGM-----YAC--DLCDK 735
 Db 286 PLSTIKAFATQSGSDLKALNTTYSQSLIKPSRMKTPNRSKTPPHERPYACPVESCDR 345
 QY 736 TFOKSSSLRKHKEHTGKRPQCOICKAFKHKHLIHSRLHSGEKEPYQCDKCGKRFSH 795
 Db 346 RFSRSDLTRHIRHTGQKPFQCRICMNFSRSDHLTHIRHTGKEKPFACDICGRKFAR 405
 QY 796 SGYSQGMNHRYSYCKRAEBEREAEREA 824
 Db 406 SDEKRRHTKIHRLROKDKKXDVASPA 434

RESULT 8

US-08-466-344-1
 ; Sequence 1, Application US/08466344
 ; Patent No. 5773583
 ; GENERAL INFORMATION:
 ; APPLICANT: Sukhatme, Vikas P.
 ; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO THE
 ; NUMBER OF SEQUENCES: 67
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White & Durkee
 ; STREET: 321 No. 5773583th Clark Street, Suite 800
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: U.S.A.
 ; ZIP: 60610
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/466,344
 ; FILING DATE: 06-JUN-1995
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/040,548
 ; FILING DATE: 31-MAR-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Coughlin, Daniel F.
 ; REGISTRATION NUMBER: 36,111
 ; REFERENCE/DOCKET NUMBER: arc067
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (312) 744-0090
 ; TELEFAX: (312) 245-4961
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 533 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-08-466-344-1

Query Match 5.5%; Score 274; DB 1; Length 533;

Best Local Similarity 25.4%; Pred. No. 2.1e-13;
Matches 99; Conservative 47; Mismatches 143; Indels 100; Gaps 14;

QY 518 SSTS-----KNSHSSSTYTPNSFSSE-----LQAEPL-DLSLPQOMEKPGIITKTK 565
DB 64 SSTS-----KNSHSSSTYTPNSFSSE-----LQAEPL-DLSLPQOMEKPGIITKTK 565
QY 566 TKATSLNLDHNSVSSSENSDEPL-----NLTFIKKEFSNNLNKNNPVFGMPPSAK 621
DB 120 SQTTRL-----PITYTGRFSLEBPAPNSGNTLWPEPLFSLVSGVSMNTNPPTSSS----- 170
QY 622 PLVTPPLPPOSAPFPATFMPVQTSIFG-----LRPYGLDQMSFLPHMAYTYPT 670
DB 171 ---APSPASSSSSSAQSPPPLSCAVSNDSPSYAAPFPPTNTD--IFPEPQSAFPG 225
QY 671 GAATFADMQORRYKQKGF-----QGDLIDGAQDY--MSGLDWTDSDSC- 714
DB 226 SAGTALQYPPAPYATKGFQVPMIDYLPQOQGLSLGTPOKPFQGLENTQPSLT 285
QY 715 -LSRKIKKTESGM-----YAC--DLCDK 735
DB 286 PLSTIKAFATQSGQDLKALNTTYQSOLIKPSRMKYPNRPSTKTPPHERPYACPVESCDR 345
QY 736 TFQKSSSLRHKYEHTGKRPPOCQICKAFKHKHILHSRLHSGEKPYQCDKCGKRRSH 795
DB 346 RFSRSDLTFRHRIHTGOKPFCRCIRMFSDHLLTHIRTHTGKRPACDICGKRRAR 405
QY 796 SCSYSQNMHNRYSYCKREAEEREAEREA 824
DB 406 SDERKHTKIHLRQKOKADKSVASPA 434

RESULT 9
5206152-2
; Patent No. 5206152
; APPLICANT: SUKATME, VIKAS P.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF EARLY GROWTH
; REGULATORY PROTEIN GENES
; NUMBER OF SEQUENCES: 18
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/249,584
; FILING DATE: 26-SEP-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 179,587
; FILING DATE: 08-APR-1988
; SEQ ID NO: 2:
; LENGTH: 533
5206152-2

Query Match 5.5%; Score 274; DB 6; Length 533;
Best Local Similarity 25.4%; Pred. No. 2.1e-13;
Matches 99; Conservative 47; Mismatches 143; Indels 100; Gaps 14;

QY 518 SSTS-----KNSHSSSTYTPNSFSSE-----LQAEPL-DLSLPQOMEKPGIITKTK 565
DB 64 SSTS-----KNSHSSSTYTPNSFSSE-----LQAEPL-DLSLPQOMEKPGIITKTK 565
QY 566 TKATSLNLDHNSVSSSENSDEPL-----NLTFIKKEFSNNLNKNNPVFGMPPSAK 621
DB 120 SQTTRL-----PITYTGRFSLEBPAPNSGNTLWPEPLFSLVSGVSMNTNPPTSSS----- 170
QY 622 PLVTPPLPPOSAPFPATFMPVQTSIFG-----LRPYGLDQMSFLPHMAYTYPT 670
DB 171 ---APSPASSSSSSAQSPPPLSCAVSNDSPSYAAPFPPTNTD--IFPEPQSAFPG 225
QY 671 GAATFADMQORRYKQKGF-----QGDLIDGAQDY--MSGLDWTDSDSC- 714
DB 226 SAGTALQYPPAPYATKGFQVPMIDYLPQOQGLSLGTPOKPFQGLENTQPSLT 285
QY 715 -LSRKIKKTESGM-----YAC--DLCDK 735
DB 286 PLSTIKAFATQSGQDLKALNTTYQSOLIKPSRMKYPNRPSTKTPPHERPYACPVESCDR 345

QY 736 TFQKSSSLRHKYEHTGKRPPOCQICKAFKHKHILHSRLHSGEKPYQCDKCGKRRSH 795
DB 346 RFSRSDLTFRHRIHTGOKPFCRCIRMFSDHLLTHIRTHTGKRPACDICGKRRAR 405
QY 796 SCSYSQNMHNRYSYCKREAEEREAEREA 824
DB 406 SDERKHTKIHLRQKOKADKSVASPA 434

RESULT 10
US-08-224-482-4
; Sequence 4, Application US/08224482
; Patent No. 5837692
; GENERAL INFORMATION:
; APPLICANT: Mercola, Dan
; APPLICANT: Adamson, Eileen D.
; TITLE OF INVENTION: Inhibition of the Mitogenic Activity of
; TITLE OF INVENTION: PDGF by Mammalian EGR
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/224,482
; FILING DATE: 07-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-ME 9913
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-224-482-4

Query Match 5.5%; Score 272.5; DB 2; Length 543;
Best Local Similarity 26.3%; Pred. No. 2.8e-13;
Matches 96; Conservative 47; Mismatches 133; Indels 89; Gaps 14;

QY 516 NLSSTNS-----KNSHSSSTYTPNSFSSE-----LQAEPL-DLSLPQOMEKPGI 559
DB 61 NLSSTNS-----KNSHSSSTYTPNSFSSE-----LQAEPL-DLSLPQOMEKPGI 559
QY 560 IYTKKTKATSLNLDHNSVSSSENSDEPL-----NLTFIKKEFSNNLNKNNPVFGM 615
DB 117 VETSPSQTTRL-----PITYTGRFSLEBPAPNSGNTLWPEPLFSLVSGVSMNTNPPTSSS 172
QY 616 ---NPSAKPLVTPPLPPOSAPFPATFMPVQTSIFGLRYPGLDQMSFLPHMAYTYPTGA 672
DB 173 SAPSPASSSSSSAQSPPPLSCAVSNDSPSYAAPFPPTNTD--IFPEPQSAFPGSA 229
QY 673 ATPADMQORRYKQKGF-----QGDLIDGAQDY--MSGLDWTDSDSC--L 715
DB 230 GYALQYPPAPYATKGFQVPMIDYLPQOQGLSLGTPOKPFQGLENTQPSLT 289
QY 716 SRSKIKKTESGM-----YAC--DLCDK 737
DB 290 STIKAFATQSGQDLKALNTTYQSOLIKPSRMKYPNRPSTKTPPHERPYACPVESCDR 349

QY 738 QKSSSLRHKYEHTGKRPQCOICKAFKHKHLLIEHSRLHSGEKPYOCDKGRFSG 797
Db 350 SRSELTTHIRIHGTQKPFQCRICMRNFSRSDHLLTTHIRHTGKPFACDICKRFA 409
QY 798 SYSQH 802
Db 410 ERKRH 414

RESULT 11

US-09-702-705-337
; Sequence 337, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaityana S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darriack
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 337
; LENGTH: 543
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-702-705-337

Query Match 5.5%; Score 272.5; DB 4; Length 543;
Best Local Similarity 26.3%; Pred. No. 2.8e-13;
Matches 96; Conservative 47; Mismatches 133; Indels 89; Gaps 14;

QY 516 NLSTSS-----KNSHSSSTYTPNSFSSE-----LQAEPL-DLSLPKQREPGI 559
Db 61 NSSSSSGGGGGGGGSGNSSSSTFNPQADTGEQPYEHLTAESFPDISLNNE-----KVL 116
QY 560 IATNKTATSIINDHNSVSSSENDEPL-----NLTFIKKEFSNNLNDKSNPNVGM 615
Db 117 VETISPGQTRL-----PPTTYTGRFSLBPAPNSGTTLMPEPLFSLVSGLVMTNPPASS 172
QY 616 ---NPFSAKPLTYPLPQSAFPATFMPVOTSIPLGRPYGLDOMSFLPHMAYTYPTGA 672
Db 173 SAPPAASASASQSPPLSCAVPSNDSPISAAPTF-PTENTD--IFPEQSQAFPSA 229
QY 673 ATPADMOORRYQKQGF-----QGLLDGAQDY--MSGIDMTSDSC--L 715
Db 230 GTALQPPPAYPAKGFQVPMIPDYLPQOQGLGTDPQKPFQGLESTQOPLPL 289
QY 716 SRKIKKTESGM-----YAC--DLCDKTF 737
Db 290 STTAPATQSGSDILKALNTSYQSLIKPSMRKYPNRPSTPPHERPYACPVESCDRRF 349
QY 738 QKSSSLRHKYEHTGKRPQCOICKAFKHKHLLIEHSRLHSGEKPYOCDKGRFSG 797
Db 350 SRSELTTHIRIHGTQKPFQCRICMRNFSRSDHLLTTHIRHTGKPFACDICKRFA 409
QY 798 SYSQH 802
Db 410 ERKRH 414

RESULT 12
US-09-736-457-337
; Sequence 337, Application US/09736457
; Patent No. 6509448

; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaityana S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darriack
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 337
; LENGTH: 543
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-736-457-337

Query Match 5.5%; Score 272.5; DB 4; Length 543;
Best Local Similarity 26.3%; Pred. No. 2.8e-13;
Matches 96; Conservative 47; Mismatches 133; Indels 89; Gaps 14;

QY 516 NLSTSS-----KNSHSSSTYTPNSFSSE-----LQAEPL-DLSLPKQREPGI 559
Db 61 NSSSSSGGGGGGGGSGNSSSSTFNPQADTGEQPYEHLTAESFPDISLNNE-----KVL 116
QY 560 IATNKTATSIINDHNSVSSSENDEPL-----NLTFIKKEFSNNLNDKSNPNVGM 615
Db 117 VETISPGQTRL-----PPTTYTGRFSLBPAPNSGTTLMPEPLFSLVSGLVMTNPPASS 172
QY 616 ---NPFSAKPLTYPLPQSAFPATFMPVOTSIPLGRPYGLDOMSFLPHMAYTYPTGA 672
Db 173 SAPPAASASASQSPPLSCAVPSNDSPISAAPTF-PTENTD--IFPEQSQAFPSA 229
QY 673 ATPADMOORRYQKQGF-----QGLLDGAQDY--MSGIDMTSDSC--L 715
Db 230 GTALQPPPAYPAKGFQVPMIPDYLPQOQGLGTDPQKPFQGLESTQOPLPL 289
QY 716 SRKIKKTESGM-----YAC--DLCDKTF 737
Db 290 STTAPATQSGSDILKALNTSYQSLIKPSMRKYPNRPSTPPHERPYACPVESCDRRF 349
QY 738 QKSSSLRHKYEHTGKRPQCOICKAFKHKHLLIEHSRLHSGEKPYOCDKGRFSG 797
Db 350 SRSELTTHIRIHGTQKPFQCRICMRNFSRSDHLLTTHIRHTGKPFACDICKRFA 409
QY 798 SYSQH 802
Db 410 ERKRH 414

RESULT 13
US-09-449-285A-21
; Sequence 21, Application US/09449285A
; Patent No. 6313280
; GENERAL INFORMATION:
; APPLICANT: Vlaams Interuniversitair Instituut voor Biotechnologie
; TITLE OF INVENTION: SMAD-INTERACTING POLYPEPTIDES AND THEIR USE
; FILE REFERENCE: 2676-4232US
; CURRENT APPLICATION NUMBER: US/09/449,285A
; CURRENT FILING DATE: 1999-11-24
; PRIOR APPLICATION NUMBER: PCT/EP98/03193
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 97201645.5
; PRIOR FILING DATE: 1997-06-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0

SEQ ID NO 21
LENGTH: 51
TYPE: PRT
ORGANISM: SBD mutant
US-09-449-285A-21

Query Match
Best Local Similarity 100.0%; Pred. No. 1e-13;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

166 OHGVGEAPLIGFPPTMNSNLSEVQKVLQIVDNTVSRQKMDCKTEDISKLG 216
1 OHGVGEAPLIGFPPTMNSNLSEVQKVLQIVDNTVSRQKMDCKTEDISKLG 51

RESULT 14
US-08-459-568-2
Sequence 2, Application US/08459568
Patent No. 5811304

GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
TITLE OF INVENTION: Zinc Finger Proteins
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,568
FILING DATE: 02-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/399,411
FILING DATE: 06-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULAR TYPE: protein
US-08-459-568-2

Query Match
Best Local Similarity 18.8%; Score 256.5; DB 2; Length 1706;
Pred. No. 3.2e-11;
Matches 200; Conservative 132; Mismatches 381; Indels 353; Gaps 41;

11 FKCTEGCAFKYKHLKEHLRIHSG--EKRYECPNCKKRF-----HSGYSYSHISK 61
355 FPCQHCERKFAFGQERHMHIIITINAFKCKYCGKFGQINRRHRHREFTGLKRR 414
62 KCIGLISVNGRKNRIKGTSSSPNSAITOLRNKLENGK-----LSMSEQT 114
445 PSMTLQS-----SEDDPGKGVNTSKDESSPQLGQDCILINSSEKT 456
115 GLLKIKTEPLDNDYKVLMAHFGSSSPFM--NGLGATSP-----LGVHPSQSPM 165
457 -----SQEVL-----NSSFVEENGKELHPCKYKVGTHITNRRHQ 495

166 -----QHL---GVGEAPLIGFPPTMNSNLSEVQKVLQIVDNTVSRQKMDCKTEDISKLG 217
496 RRVHERHLIPKGVRRKGLIEEPQPAEQAPPSQVYV----- 533
218 YHMKPCSQPEEQGYT-----SPNIPVQGLPVVSHNGATKS----- 253
534 -----PSTPEEEGETDDVYIMDISNISENLNYYIDGKIQTNSSTNSCNDVIEMESNASH 588
254 -----IIDYLEKVENEAKACLOSLTTDSRR---QISNIKEKRLTLLIDLVTDKMIEN 303
589 LYGIDCLLPVVEITONIKSTQVAVTDLDLDSBSTCSCKRRTAPPLPKITET 648
304 HSIISTPFCQCKESFPFGPIPLHQ---HERYLCKNNEIKAVLQPHENI----- 349
649 ESDSTAPSCSL---SLPLSISTAEVVSFFKEKGVLYSLKLOLQOTDCLTLPAGFSAE 705
350 VENKAGVFDNKAALL---SSVLSKGLSPINPKDHSVLKAYAMMEEDSLKAI 406
706 IPLGPGVCSADPASMPLVPTSSRPKRRTSSPPSP--QHSPLALDF----- 748
407 SIAGVLPQEFVKEWEQKRVQYSNRSPLERTSKPLAPNSPTTKOSILPRSVPKMD 466
749 -----GKPNCKAKAMD-----IVLTSKPKLB--SRSDPAMSLSGRDERETGSP----- 792
467 SITSPSIAELHNSVTSQDPLRLTK---SSHPTNIKAVDKLDHSKNTSPPLNLSSTSS 522
793 -----PCFDEYKISKEMAASSTFSSV-----CNOQLDLSSGVK 826
523 KNSHSSYTPNFSSELEQAEPLDLSLPRKMEPGIATKN-----KTKATSNL- 573
827 QKSEGTGKTPVWES-----VLDSLHKKPCDSSCKEKENHLAQPAKKKKPTTCMLQ 880
574 -----DHSVSSSENSEDEPLNLTETIKKFSNNLNDK-----SNNPVF 613
881 KYLANEYNVSLPTETTP-----VTRSPSCSKPDTPQDELGPDSQSVPTASPPBE 934
614 GNMPFS-----AKPLYTPLPOS-----APPATMPPVQTSIPGL- 649
935 VVGPSSPPLQATSLSSGQLPRLTPTPEPSPPCPPVLTVATPPPLPLTVPLSHSPSDA 994
650 -----RPYRGLDQMSFLPHMAVTP-----TGAATFADMQORRYQRQGF- 690
995 SPQOCSPSPSNTTAOSPLILSPVSPSPPIPVPEPLMASPGPPTLSSSSSSSSSP 1054
691 -----QGDLLDGA-----QDYMSGIDMDTSDSCISRRK 719
1055 SSSCSTSPSPPLAVSSVSSGDNLEASLPAVTFKQESSEBGLKPEEAPRPGQGV 1114
720 IKKTESGMYACDLCDKTFQKSSSLRHKYEHTGKRPQOCICKAFKXHLIENS-RLH 778
1115 VOETFSKNFICNVCSPPFLSIDLTGKLSVHAEEWPFKCFVQLFKVTDLSEHRLFLH 1174
779 SGEKPYQCDKCGKRRFSHSGSYSGHNMHRYSQKRAEAEAREAREKRGHGLPPELLMN 838
1175 GVGNIFVCSVCKKEPAFLCNLQOH-----QRDLHPDEVCTHHFESGTLP----- 1220
839 RAYLOSITPQGYSDSEERESMPRDG-ESEKEHE--KEGEGYGKLR 881
1221 -----QNFIDPSKANYEHMSLPEEPLTISRREBLNDSSELYTTIK 1262

RESULT 15
US-08-399-411-2
Sequence 2, Application US/08399411
Patent No. 5811008
GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
TITLE OF INVENTION: Zinc Finger Proteins
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/399,411
FILING DATE: 06-MAR-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-399-411-2

Query Match 5.1%; Score 256.5; DB 2; Length 1706;
Best Local Similarity 18.8%; Pred. No. 3,2e-11;
Matches 200; Conservative 132; Mismatches 381; Indels 353; Gaps 41;

11 FKCTCGAFKFKYKHLKHLRHS--EKYECFPCNCKRFS-----HSGSYSHISK 61
355 PFCQCEKFKATQGLERHMHISTINAFKCKYCGKRFQIRRRHERHETGLKR 414
62 KCIQLISVNGMRNNIKTGSSPNSVSSPTNSAITOLRNKLENGK-----LSMSBOT 114
415 PSMTLQS-----SEDDDGKGNVTSKDESSPQLODCLINSEKT 456
115 GLKTKTEPLDNDYKVLMTATGFGSGSPFM--NGGLGATSP-----LGVHPAQSFM 165
457 -----SOEVL-----NSFVEENGVEKELHPCKYCKVFGTHMRRHQ 495
166 -----OHL-----GVGMEAPLPGPTMNSLSVQXVLQIVDNTVSRQKMDCKTEDISKLG 217
496 RRVHERHLIPKGVRRKGLLEBPAPAOAPPSQNVV-----533
218 YHMKDPCSQPEQGV-----SPNIPVGLPVSHNGATKS-----253
534 -----PSTEPEEGEFTDVIYIMDISNISNENIYIDGKIOTNSSTNSCDVIEMSNSAH 588
254 -----IIVTYLKEVNEAKACLOSLTDSRR--OISNIKKEKRLTLDVTDKMIEN 303
589 LYGIDCLLTPTVVEITONIKSTOVSVTDLLKDSPTSCSKRRITASPPVLPIKTIET 648
304 HSIPTFCQCFKESFPGPIPLHQ-----HERYLCQNEEIKAVIQPHENI-----349
649 ESDSTAPSCST--SLPLISTAEVVSFHKGVYLISSKQLQLODCLTPAGFSAAE 705
350 VPKAGVFDVNDKALL-----SSVLSSEKGLTSPINPYKDHMSVLKAYVAMNMBNSDELKI 406
706 IPKLGVCASAPASMLPTVTSRFRKRTSSPSSP--QHSPLALDF-----748
407 SIANGLPQEFVKEWFOQRKVYQYSNSRSPSLERTSKPLAPNSPTTKDLSLPRSPVKPM 466
749 -----GKPNDGKAATD-----TVLTSKKPKLE--SRSDSPAWLSGRDERETGSP-----792
467 SITSPSIMEIHSVSCDPLRLTK-----SSHFTNIKAVDKLDHRSNTPSPPLNISSTSS 522
793 -----PCFDEYKISKEMASTFSSV-----CNQOPLDLSGKV 826

523 KNSHSSSTPNSFSSEELQAEPLDISLPKQNRKPKGIATON-----XTKATSNLV 573
827 QKSBGTGKTPVPMWS-----VLDSLVRHKPCDSEKGEFKNHILAQPAKKKKPTTCMLQ 880
574 -----DHNSVSSSENDSPELNTLTFIKKEFSNNDLKK-----SNNVVF 613
881 KYLLNEVGVSLPTETPE-----VTRSPCKSPDTQPPPELGPDSSCVPTAESPE 934
614 GNNPFS-----AKPLYPLPPOS-----APPATFMPVQTSIPGL-- 649
935 VVGPSPPLOQNASLSSGQLPPLITPTESSPPPCPPVLTVAITPPPLPTVPLSHSPSSDA 994
650 -----RPYGLDQMSFLPHAYTP-----TGATPADWQORRYQKQGF- 690
995 SPQCPSPFSNTTAQSPPLILSPVSPSPDPVPEPLMSAASPPPTLSSSSSSSSSFP 1054
691 -----QGDLDGA-----QDMSGIDMTSDSLSRK 719
1055 SSSCGSTSPSPPLSAVSSVSSGDNLEASLPAYTFKQEESESEGLKPKKEAPPAAGQSV 1114
720 IKTESGMYACDLCDKTFQKSSSLLRHKYBHTGKRPQCQICAKFKKHLIHS--RLH 778
1115 VQETFSKPIGNVCSPPPLSLKDLTKHLSVAEWPFCFCVQLFKYKTDLSHRFLH 1174
779 SGEKPYQCKCKGKRPSSHSGSYSQHMNRYSYCKEAEEREAERAREKHLGPTTELLMN 838
1175 GVGNIFVCSVCKKEPAFLCNLQOH-----QRDLHPDEVCTHHEFESGTLRP-----1220
839 RAYLOSITPQGSYSDSEKESMPDGC--ESEKEHE--KEESEGYGKLR 881
1221 -----QNFTPSKANVEHMPSLPEEPLETREBELNDSSEELVYTIK 1262

Search completed: January 5, 2004, 16:36:09
Job time : 25 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2004, 16:31:07 ; Search time 38 Seconds
(without alignments)
4984.350 Million cell updates/sec

Title: US-09-964-238-2

Perfect score: 4991
Sequence: 1 MLTQAGNRKFKCTECGKAF.....DGKMKTSDEHEDNEDMG 944

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4991	100.0	944	9 US-09-964-238-2	Sequence 2, Appli
2	2612	52.3	533	9 US-09-864-761-37151	Sequence 37151, A
3	2612	52.3	533	12 US-10-029-386-33350	Sequence 33350, A
4	1647.5	33.2	1125	10 US-09-974-298-114	Sequence 114, Ap
5	1647.5	33.0	1104	12 US-10-104-047-2506	Sequence 2506, Ap
6	1606.5	32.2	1154	15 US-10-233-561-12	Sequence 10, Appl
7	1566	31.4	1117	15 US-10-233-561-12	Sequence 12, Appl
8	1505	20.1	190	9 US-09-864-761-34862	Sequence 34862, A
9	1505	20.1	190	9 US-09-864-761-34862	Sequence 34862, A
10	356.5	7.1	1042	15 US-10-097-340-83	Sequence 83, Appl
11	355.5	7.0	1051	15 US-10-097-340-83	Sequence 79, Appl
12	349.5	7.0	1204	12 US-10-029-386-33060	Sequence 33060, A
13	329.5	6.6	269	9 US-09-864-761-38036	Sequence 38036, A
14	329.5	6.6	269	12 US-10-029-386-33074	Sequence 33074, A
15	318	6.4	1357	12 US-10-295-027-1199	Sequence 1199, Ap

16	299.5	6.0	924	15 US-10-007-280A-231	Sequence 231, App
17	297	6.0	610	10 US-10-094-749-2604	Sequence 2604, Ap
18	296.5	5.9	610	10 US-09-989-920-212	Sequence 212, App
19	294	5.9	648	12 US-10-094-749-2472	Sequence 2472, Ap
20	290	5.8	514	12 US-10-029-386-34112	Sequence 34112, A
21	282.5	5.7	576	12 US-10-094-749-2365	Sequence 2365, Ap
22	279	5.6	902	12 US-10-029-386-32110	Sequence 32110, A
23	278.5	5.6	742	12 US-10-029-386-31985	Sequence 31985, A
24	276.5	5.5	706	9 US-09-107-058-2	Sequence 2, Appli
25	276.5	5.5	706	9 US-09-761-117-2	Sequence 2, Appli
26	276	5.5	816	12 US-10-029-386-33143	Sequence 33143, Ap
27	275	5.5	545	12 US-10-144-194A-26	Sequence 26, Appl
28	275	5.5	811	12 US-10-144-194A-26	Sequence 98, Appl
29	275	5.5	811	12 US-10-144-194A-26	Sequence 36148, A
30	273.5	5.5	536	9 US-09-864-761-36148	Sequence 81, Appl
31	273.5	5.5	727	15 US-10-097-340-81	Sequence 337, App
32	272.5	5.5	543	10 US-09-736-457-337	Sequence 337, App
33	272.5	5.5	543	10 US-09-902-941-337	Sequence 337, App
34	272.5	5.5	543	10 US-09-849-626-337	Sequence 337, App
35	272.5	5.5	543	11 US-09-476-100-337	Sequence 6, Appli
36	272.5	5.5	543	12 US-10-251-661-6	Sequence 337, App
37	272.5	5.5	543	12 US-10-113-872-337	Sequence 180, App
38	272.5	5.5	543	12 US-10-116-275-180	Sequence 184, App
39	272.5	5.5	543	12 US-10-341-434-184	Sequence 184, App
40	272.5	5.5	543	12 US-10-341-434-198	Sequence 337, App
41	272.5	5.5	543	15 US-10-017-754-337	Sequence 95, Appl
42	272.5	5.5	543	15 US-10-205-823-95	Sequence 43107, A
43	270.5	5.4	614	9 US-09-864-761-43107	Sequence 11, Appl
44	269.5	5.4	566	12 US-10-322-774-10	Sequence 77, Appl
45	269.5	5.4	718	15 US-10-097-340-77	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-09-964-238-2
Sequence 2, Application US/09964238
Patent No. US20020035246A1
GENERAL INFORMATION:
APPLICANT: Vlaams Interuniversitair Instituut voor Biotechnologie
TITLE OF INVENTION: SWAD-INTERACTING POLYPEPTIDES AND THEIR USE
FILE REFERENCE: 2676-4232US
CURRENT APPLICATION NUMBER: US/09/964,238
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 09/449,285
PRIOR FILING DATE: 1999-11-24
PRIOR APPLICATION NUMBER: PCT/EP98/03193
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 97201645.5
PRIOR FILING DATE: 1997-06-02
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 944
TYPE: PRT
ORGANISM: Mus musculus
US-09-964-238-2

Query Match 100.0%; Score 4991; DB 9; Length 944;
Best Local Similarity 100.0%; Pred. No. 3.36-284;
Matches 944; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLTQAGNRKFKCTECGKAFKKKHLKHLRHSGEKYECNCCKRRSHSSSYSSHS	60
DB	1	MLTQAGNRKFKCTECGKAFKKKHLKHLRHSGEKYECNCCKRRSHSSSYSSHS	60
QY	61	KKICGLIVNGRMNRNITGSSPNVSSPTNSATTOIRNKLKNGKPLSMSEOTGLKIK	120
DB	61	KKICGLIVNGRMNRNITGSSPNVSSPTNSATTOIRNKLKNGKPLSMSEOTGLKIK	120
QY	121	TEPLDFNDYKVLMTATGFGSSPFNNGGLGATSPGLVHPASQSPMOHLGVGMEADLGGP	180

```

Db 121 TEPDLNDYKVLMTATGSSSSPPMNGGLATSPDLGVHSAQSPMOHLGVMEAPLLIGFP 180
Qy 181 TMSNLSSEVQKVLQIYDNTVSRQKMDCTEDISKAGYHMKDCSQPEEGVTSPIPIPV 240
Db 181 TMSNLSSEVQKVLQIYDNTVSRQKMDCTEDISKAGYHMKDCSQPEEGVTSPIPIPV 240
Qy 241 GLPVSHNGATKSIIDYTLERKNEAKACLOSLTDSRRQISNIKKEKRLTIDLVTDDK 300
Db 241 GLPVSHNGATKSIIDYTLERKNEAKACLOSLTDSRRQISNIKKEKRLTIDLVTDDK 300
Qy 301 IENHSITPSCQCFCKESFPGPILQHHERYLCKMNEIKAVIOPHNIPYKAGYVDN 360
Db 301 IENHSITPSCQCFCKESFPGPILQHHERYLCKMNEIKAVIOPHNIPYKAGYVDN 360
Qy 361 KALLSSVLSSEKGLTSPINPKYKMSYLKAYAMNENPSDELKISIAVGLPOEFYK 420
Db 361 KALLSSVLSSEKGLTSPINPKYKMSYLKAYAMNENPSDELKISIAVGLPOEFYK 420
Qy 421 FEORRYQVSNSSPSLERTSKPLAPNSPTTQSLIPSPVYKPMDSITSPSIAELHNSV 480
Db 421 FEORRYQVSNSSPSLERTSKPLAPNSPTTQSLIPSPVYKPMDSITSPSIAELHNSV 480
Qy 481 TSCDPLRLTKSHFTNFKAVDKLDHRSNTSPPLNLSSTSSKNSHSSVTPNSFSEEL 540
Db 481 TSCDPLRLTKSHFTNFKAVDKLDHRSNTSPPLNLSSTSSKNSHSSVTPNSFSEEL 540
Qy 541 QAEPLDLSLPKQREPKGIATKTKATISINLDHNSVSSSSSENDEPLNLTPIKKEFSN 600
Db 541 QAEPLDLSLPKQREPKGIATKTKATISINLDHNSVSSSSSENDEPLNLTPIKKEFSN 600
Qy 601 SNNLDKSNNPVGMNPFSAKPLYTPLPPOSAPPATMPVPVOTSTIGLAPYGLDMSF 660
Db 601 SNNLDKSNNPVGMNPFSAKPLYTPLPPOSAPPATMPVPVOTSTIGLAPYGLDMSF 660
Qy 661 LPHMATYPTGATFPADMQRKRYKQKQFQGLDLSAOLYSGLDMDMTSDSCLSKKI 720
Db 661 LPHMATYPTGATFPADMQRKRYKQKQFQGLDLSAOLYSGLDMDMTSDSCLSKKI 720
Qy 721 KKTESGMYACDLCTFKQSSSLRNHYETGRPHQCOI CKKAFKIHILIHSLRHSG 780
Db 721 KKTESGMYACDLCTFKQSSSLRNHYETGRPHQCOI CKKAFKIHILIHSLRHSG 780
Qy 781 EKYEQCDCKGRSHSGSYQHMNHRSYCKREAEEREAERAREKGLGPIELLMNRA 840
Db 781 EKYEQCDCKGRSHSGSYQHMNHRSYCKREAEEREAERAREKGLGPIELLMNRA 840
Qy 841 YLOSITPQGSDBEEREMPRDSEKHEKEGEGYGLRRDGDSEEESESESNKS 900
Db 841 YLOSITPQGSDBEEREMPRDSEKHEKEGEGYGLRRDGDSEEESESESNKS 900
Qy 901 MDTDPETIRDEEFTGDHSDSDSGKMETKSDHEEDNMDGNG 944
Db 901 MDTDPETIRDEEFTGDHSDSDSGKMETKSDHEEDNMDGNG 944

```

RESULT 2

```

; Sequence 37151, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26

```

```

; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37151
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010130.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
; OTHER INFORMATION: EST HUMAN HIT: A1879705.1, EVALUE 3.00e-87
; OTHER INFORMATION: SWISSPROT HIT: P37275, EVALUE 2.00e-45
; US-09-864-761-37151

```

Query Match 52.3%; Score 2612; DB 9; Length 533;

Best Local Similarity 94.2%; Pred. No. 4e-145;

Matches 502; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

```

Qy 36 EKYEQCDCKGRSHSGSYSHSSKCTGLISVNGRMNNIKTGSSPNSVSSPTNSAI 95
Db 1 EKYEQCDCKGRSHSGSYSHSSKCTGLISVNGRMNNIKTGSSPNSVSSPTNSAI 60
Qy 96 TOLRNLKNGKPLSMSEOTGLIKITEPLDFNDYKVLMTATGSSSPMNGGLGATSP 155
Db 61 TOLRNLKNGKPLSMSEOTGLIKITEPLDFNDYKVLMTATGSSSPMNGGLGATSP 120
Qy 156 GVHSAQSPMOHLGVMEAPLLGFPMTNSNLSEVQKVLQIYDNTVSRQKMDCTEDISK 215
Db 121 GVHSAQSPMOHLGVMEAPLLGFPMTNSNLSEVQKVLQIYDNTVSRQKMDCAEISKL 180
Qy 216 KYTHMDPCSQPEEGVTSPIPIPVGLPVVSHNGATKSIIDYTLERKNEAKACLOSLTD 275
Db 181 KYTHMDPCSQPEEGVTSPIPIPVGLPVVSHNGATKSIIDYTLERKNEAKACLOSLTD 240
Qy 276 SRRQISNIKKEKRLTIDLVTDDKMIENHSITPSCQCFCKESFPGPILQHHERYLCKM 335
Db 241 SRRQISNIKKEKRLTIDLVTDDKMIENHSITPSCQCFCKESFPGPILQHHERYLCKM 300

```

QY 336 NEEIAVLOPHENIYVKNAGVVDNKALLLSVLSSEKGLTSPINPKYKDHMSVTKAYYAN 395
DB 301 NEEIAVLOPHENIYVKNAGVVDNKALLLSVLSSEKGLTSPINPKYKDHMSVTKAYYAN 360
QY 396 MEPSDELKLSIAVGLPOEFVKEMFEORKYQYNSNSPSLERSKPLAPNSNPTTDS 455
DB 361 MEPSDELKLSIAVGLPOEFVKEMFEORKYQYNSNSPSLERSKPLAPNSNPTTDS 420
QY 456 LLPSPVPMDSITSPSIAELHNSVTSQDPLRLTKSSHFTNIKAVDKLDSRSNTPSPL 515
DB 421 LLPSPVPMDSITSPSIAELHNSVTSQDPLRLTKSSHFTNIKAVDKLDSRSNTPSPL 480
QY 516 NLSSTSSKNSSSSSYTPNSFSSSEELQAEFFRLVITKTERTOKYSHKEQNK 568
DB 481 NLSSTSSKNSSSSSYTPNSFSSSEELQAEFFRLVITKTERTOKYSHKEQNK 533

RESULT 3

US-10-029-386-33350
; Sequence 33350, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33350
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010130.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.45
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
; OTHER INFORMATION: SWISSPROT HIT: O60315, EVALU0 0.00e+00
US-10-029-386-33350

Query Match 52.3%; Score 2612; DB 12; Length 533;
Best Local Similarity 94.2%; Pred. No. 4e-145;

Matches 502; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

QY 36 EKPVECPNCKRFSHSGSYSHISSKCGILISVNGRMNNIKTSSPNSVSSSPTNSAI 95
DB 1 EKPVECPNCKRFSHSGSYSHISSKCGILISVNGRMNNIKTSSPNSVSSSPTNSAI 60
QY 96 TOLARKLNGKRLNSSEOTGLIKITEPLDNDYVLAATHGSSSPMNGGLATSP 155
DB 61 TOLARKLNGKRLNSSEOTGLIKITEPLDNDYVLAATHGSSSPMNGGLATSP 120
QY 156 GVHPSAOSPMOHLGVGMEAPLGFPTMNSNLSVOVKVLOIVNTVSRQMDCKTEIDISL 215
DB 121 GVHPSAOSPMOHLGVGMEAPLGFPTMNSNLSVOVKVLOIVNTVSRQMDCKTEIDISL 180
QY 216 KGYHMKDPCSOPBEQGVTSFNI.PVGLPVVSHNGATKSIIDYTLKVNKAACLOSITTD 275
DB 181 KGYHMKDPCSOPBEQGVTSFNI.PVGLPVVSHNGATKSIIDYTLKVNKAACLOSITTD 240
QY 276 SRRQISNKKKELRLTLDVTDKMIENHSISTPSCQCFKSPGPIPLHGERYLCM 335
DB 241 SRRQISNKKKELRLTLDVTDKMIENHSISTPSCQCFKSPGPIPLHGERYLCM 300
QY 336 NEEIAVLOPHENIYVKNAGVVDNKALLLSVLSSEKGLTSPINPKYKDHMSVTKAYYAN 395
DB 301 NEEIAVLOPHENIYVKNAGVVDNKALLLSVLSSEKGLTSPINPKYKDHMSVTKAYYAN 360
QY 396 MEPSDELKLSIAVGLPOEFVKEMFEORKYQYNSNSPSLERSKPLAPNSNPTTDS 455
DB 361 MEPSDELKLSIAVGLPOEFVKEMFEORKYQYNSNSPSLERSKPLAPNSNPTTDS 420

DB 361 MEPSDELKLSIAVGLPOEFVKEMFEORKYQYNSNSPSLERSKPLAPNSNPTTDS 420
QY 456 LLPSPVPMDSITSPSIAELHNSVTSQDPLRLTKSSHFTNIKAVDKLDSRSNTPSPL 515
DB 421 LLPSPVPMDSITSPSIAELHNSVTSQDPLRLTKSSHFTNIKAVDKLDSRSNTPSPL 480
QY 516 NLSSTSSKNSSSSSYTPNSFSSSEELQAEFFRLVITKTERTOKYSHKEQNK 568
DB 481 NLSSTSSKNSSSSSYTPNSFSSSEELQAEFFRLVITKTERTOKYSHKEQNK 533

RESULT 4

US-09-974-298-114
; Sequence 114, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Huel-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 114
; LENGTH: 1125
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID NO. US20020156263A1 3267030CD1
US-09-974-298-114

Query Match 33.2%; Score 1657.5; DB 10; Length 1125;
Best Local Similarity 41.9%; Pred. No. 8.9e-89;

Matches 406; Conservative 147; Mismatches 308; Indels 107; Gaps 32;

QY 2 LTQAGNRKFKCTGCAFKYKHLKHELRTHSGEPCNCKRFSHSGSYSHISSK 61
DB 232 VTQSGCNKFKCTGCAFKYKHLKHELRTHSGEPCNCKRFSHSGSYSHISSK 291
QY 62 KCIGLISVNGRMNNIKTSSPNSVSSSPTNSAITOLARKLNGKRLNSSEOTGLIKI 119
DB 292 KCISILIPNKGPRGLTKSSQCSSP-SLSASBSPRRPQIRQKIER-KL--QEOUSVNOI 347
QY 120 KTEPLDNDYVLAATHGSSSPMNGGLATSPDGVHPSAOSPMOHL---GVGMEAPL 176
DB 348 KTEPLDNDYVLAATHGSSSPMNGGLATSPDGVHPSAOSPMOHL---GVGMEAPL 176
QY 177 LGFTMNSNLSVOVKVLOI-VDNVYSRQMDCKTEIDISLKGYNMKDPCSOPBEQGVTSF 235
DB 407 ---SINLSIDIQVLAADVAGNIVROVLE---NNQNLASSKQETINASPLOQGGHSHV 457
QY 236 NIPVGLPVVSHNGATKSIIDYTLKVNKAACLOSITTDNRROISNKKKELRLTLDV 295
DB 458 -ISAISLPLVODGDTTKIILNYSLEQSQLOVVPQNLKKKPVATNSCKSKLPBDDLTVK 516
QY 296 TD-DKMIENH-SISTPSCQCFKSPGPI-PLHGERYLCRMEELIAVLOPHENIYV 352
DB 517 SEKDSFEGGVNDSTCLLDCDC---PDINALDELKHY-----DLKQPTQP-----PP 561
QY 353 KAGVVDNKALLLSVLSSEKGLTSPINPKYKDHMSVTKAYYANMEPSDELKLSIAVGL 412
DB 562 LPAAEAEKPESSVSSATDGNLSBQPLKXLLSLKAYVALNKOQSAEELSKTADSVNL 621
QY 413 POEFVKEMFEORKYQYNSNSPSLERSKPLAPNSNPTTDSLLPSPVPMDSITSPS 472
DB 622 PLDVKKKMFEKQAGQIS-----VQSSBSPSPGKV-----NTPAKNNOQPOGAN 667
QY 473 IAEHLNSVTSQDPLRLTKSSHFTNIKAVDKLDSRSNTPSPLNLSSTSSKNSSSSSYTP 532
DB 473 IAEHLNSVTSQDPLRLTKSSHFTNIKAVDKLDSRSNTPSPLNLSSTSSKNSSSSSYTP 532

```

Db      668 ANEPDSTVNLOSPLKMTNS---PVLPGVSTNGSRSSSTPSPPLNLISSSRNTQGYLYTA 724
Qy      533 NSPSSSELOAPPLDLISPKOMREPCKGIATKNTKATSIINLDHNSVSSSENS-----DE 587
Db      725 EG-AQEEPOVEPPLDLISPKQGG-----LLERSTITSYQNSVYSQVE 767
Qy      588 PLNLITIKKEFSNNLDKNSNNPVFGMNPFAKPLTYLPPOSAPPATFMPPVQTSIP 647
Db      768 PLNLISCAKKEPKQKSCVTD--SEPVNVVIPPSSANPIAIPTVTAQLPTVTAIADQNSVP 825
Qy      648 GLRPYGLDQMSFLPMATYTPPGAATFADMQRKRYQKQFGQGLDLGADQYMSGLD 707
Db      826 CLRALANKQITLIPQVAYTSTTVSP-AVQEPPLKVIOPNQWDERQDTSSEGVNSVD 864
Qy      708 MTDSDCLSRKKIKTESGMYACDLCDKTFQKSSSLRHKYHTGKRPQCOICKAFAK 767
Db      885 QNDSSTPPKCKKRTKENGMYACDLCKIFQKSSSLRHKYHTGKRPHECGICKAFAK 944
Qy      768 KHLIHSRLHSGEKPYQCKGCRFSHSGSYQHNHRYSYCKREAEEREAERAREK 827
Db      945 KHLIHSRLHSGEKPYQCKGCRFSHSGSYQHNHRYSYCKREAEERDSTEOR---- 1000
Qy      828 GHIGPELMLNRAVLOSTIP-OGYSDSEERESMPR--DGESEKHEKEGE-----GYK 879
Db      1001 -EAGP-EILSNEHVGAASPSQG--DSDERESLIREDEDESEKEBEDKEMELOEKE 1056
Qy      880 LRRDGDDEEBEESSEKNSMDTDEPTIRDEEETGDHSMDSSED-----GKMET 930
Db      1057 CEKPGQDEEBEGBEVEEVEE-EAENGEAKTEGLMKDRAESQASSLGQKVESSE 1115
Qy      931 KSDHEEDN 938
Db      1116 QVSEKTN 1123

RESULT 5
US-10-104-047-2506
; Sequence 2506, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cdna
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 2506
; LENGTH: 1104
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2506

```

```

Query Match      33.0%; Score 1647.5; DB 12; Length 1104;
Best Local Similarity 41.8%; Pred. No. 3,4e-88;
Matches 405; Conservative 146; Mismatches 310; Indels 107; Gaps 32;

```

```

Qy      2 LITGAGNRKKCTECGCAFKYKHLKHLHLSGKEYECPNCKKRPSSHSGSYSHSSK 61
Db      211 VTQSGCNRRKCTECGCAFKYKHLKHLHLSGKEYECPNCKKRPSSHSGSYSHSSK 270
Qy      62 KCIGLISVNGRMNNIKTG--SSPNVSASSPTNSAITQLNKLNGKPLMSBOTGLTKI 119
Db      271 KCISSLIPVNGRPTGLTKTSCSSP-SISASPSGPTRPQIKQIKEN-KPL--QEQLSVNOI 326
Qy      120 KTEPLDFNDYKVMATHGFSGSPFMNGLGATSPGLVHPSAOSPMQHT--GVGMEAPL 176
Db      327 KTEPVDY-EFKPIVVASGINSCTPLQNGVPTGGPGLQATSSPGQWQAVVLPVGLVSP 385
Qy      177 LGPTMANSLSVQKYLQI-VDNITVSRQKMDCTEDISKLGKGYHMDPGSQPEEGYTSP 235

```

```

Db      386 -----SINLSDIQNVLTAKVADGNVIRQYLE--NNQANLAKSEGTITNAPISOQGHV 436
Qy      236 NIPPVGLPVVSHNGATKSIIDYTLKEVNEAKACLOSLTDSRROISNIKEKRLTLDLV 295
Db      437 -ISAILPLVDQGTTKLIINYSLEBPQQLQVVPQULKENPVATNCSSEKLPEDLYK 495
Qy      296 TD-DKMIENH-SISTPSCQFCESFPPI-PLHQHERLYCKMNEIKAVLOPHEIVRN 352
Db      496 SEKDKSFEGGVNDSTCLDDCC--PGDINALPELKH-----DLKQPTQ-----PP 540
Qy      353 KAGVPVNDKALLISVLSSEKGLTSPINPKDHMSVKAAYAMMBEBSLAKISIAVG 412
Db      541 LPAAEAKESBSVSAATGNGNLSPOPLKUNLUSLKAAYALNAQSAEELKIASVNI 600
Qy      413 POFVMEFQORRVYQYNSRSPSLERTSKPLAPNSNPTTKSLPRSPVKPMDSITSPS 472
Db      601 PLDVVAKWFEKMQAGIS-----YQSEPESSPEKGV-----NIPAKNNDQPSAN 646
Qy      473 IAEHNSVTSQDPLRLTSSHTNIKAVDKLDHSNTPSPNLASTSSKSHSSYTP 532
Db      647 ANEPDSTVNLOSPLKMTNS---PVLPGVSTNGSRSSSTPSPPLNLISSSRNTQGYLYTA 703
Qy      533 NSPSSSELOAPPLDLISPKOMREPCKGIATKNTKATSIINLDHNSVSSSENS-----DE 587
Db      704 EG-AQEEPOVEPPLDLISPKQGG-----LLERSTITSYQNSVYSQVE 746
Qy      588 PLNLITIKKEFSNNLDKNSNNPVFGMNPFAKPLTYLPPOSAPPATFMPPVQTSIP 647
Db      747 PLNLISCAKKEPKQKSCVTD--SEPVNVVIPPSSANPIAIPTVTAQLPTVTAIADQNSVP 804
Qy      648 GLRPYGLDQMSFLPMATYTPPGAATFADMQRKRYQKQFGQGLDLGADQYMSGLD 707
Db      805 CLRALANKQITLIPQVAYTSTTVSP-AVQEPPLKVIOPNQWDERQDTSSEGVNSVD 863
Qy      708 MTDSDCLSRKKIKTESGMYACDLCDKTFQKSSSLRHKYHTGKRPQCOICKAFAK 767
Db      864 QNDSSTPPKCKKRTKENGMYACDLCKIFQKSSSLRHKYHTGKRPHECGICKAFAK 923
Qy      768 KHLIHSRLHSGEKPYQCKGCRFSHSGSYQHNHRYSYCKREAEEREAERAREK 827
Db      924 KHLIHSRLHSGEKPYQCKGCRFSHSGSYQHNHRYSYCKREAEERDSTEOR----- 979
Qy      828 GHIGPELMLNRAVLOSTIP-OGYSDSEERESMPR--DGESEKHEKEGE-----GYK 879
Db      980 -EAGP-EILSNEHVGAASPSQG--DSDERESLIREDEDESEKEBEDKEMELOEKE 1035
Qy      880 LRRDGDDEEBEESSEKNSMDTDEPTIRDEEETGDHSMDSSED-----GKMET 930
Db      1036 CEKPGQDEEBEGBEVEEVEE-EAENGEAKTEGLMKDRAESQASSLGQKVESSE 1094
Qy      931 KSDHEEDN 938
Db      1095 QVSEKTN 1102

```

```

RESULT 6
US-10-232-561-10
; Sequence 10, Application US/10232561
; Publication No. US20030119772A1
; GENERAL INFORMATION:
; APPLICANT: Genetex, Thomas
; TITLE OF INVENTION: Methods and compositions useful for
; FILE REFERENCE: CHOP 00-99
; CURRENT APPLICATION NUMBER: US/10/232,561
; PRIOR FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 60/317,300
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 1154
; TYPE: PRT

```

ORGANISM: Human
US-10-232-561-10

Query Match 32.2%; Score 1606.5; DB 15; Length 1154;

Best Local Similarity 41.1%; Pred. No. 9e-86; Matches 399; Conservative 142; Mismatches 314; Indels 115; Gaps 32;

```

QY 2 LTOGAGNRKFKCTGCGAKFKYKHHLKEHLRIHSGEKPEYECNCKKRFSGSHSISK 61
DB 265 VTOSGCKNRKFKCTGCGAKFKYKHHLKEHLRIHSGEKPEYECNCKKRFSGSHSISK 324
QY 62 KCIIGLISVNGRMRNNIKTG--SSPNSVSSPTNSAITQIRKLENGKPLMSBOGLIKI 119
DB 325 KCIISLIPVNGRERTIKTSQCSSP-SLSASPSGSPRPQIRKIEV-KPL--QEOULSVQI 380
QY 120 KTEPLDFNDYKVLMTATHGFGSSGSPMNGGLATSPGLGHPASQSMOH---GVMEAPL 176
DB 381 KTEPLVDY-EFKPIVVASGINSCTPLQNGVFTGGGFLQATSSPQGVAVLFTVLVSPI 439
QY 177 LGFPTMNSLSEVQKVLQI-VDNTVSRQKMDCKTEDISLKGYHMKDPCSQPEOGVTSP 235
DB 440 -----SINLSDIQVNLKAVDGNVIRQVLE--NNQANLASKEDETINASPIQGGHSV 490
QY 236 NIPRVLGVVSHNGATKSIIDYTELEKNEAKACLOSITTSRQISNKKERKLTLDLV 295
DB 491 -ISATSLPLVDQDGTTKIINYSLEQPSQLQVVPQNLKKNPVATNSCKSEKLPEDLTVK 549
QY 266 TD-DKMIENH-SISPTFGCOFCESFPGPI-PLHQHERYLCMNEIRKAVIOPHENIYPN 352
DB 550 SEKDSFEGGVNDSTCLLCCDC---PGDINALPELKH-----DLQPTP-----PP 594
QY 353 KAGVFNVDKALLLSVLESEKGLTSPINPYKDMSVLKAYYANMNEPNSDELKISAVGL 412
DB 595 LPAEAEKPESSVSSATGDNLSPOQPLKXNLISLKKAYALNAPSAEELSKINDSVNL 654
QY 413 POEFYKEMFEQKRYQYNSRSPLERTSKPLAPNSNPTTKOSLIPRPVKPMDSITSPS 472
DB 655 PLDVKKMKFEKMOAQOIS-----VQSEPESSPEPGKV-----NTPAKNNDPOQSAN 700
QY 473 IAEHLNSTVSCDPLRLTKSSHFTNIKAVDKLDRSMSTPPLNLSSTSKSHSSSTTP 532
DB 701 ANEPDSTVTNLOSPLKMTNS---PVLPGSTTNGSRSSPTSPSLNLSRRTQGYLYTA 757
QY 533 NSFSEELQAEPLDLSLPKOME-----PKGIATKTKATKATINLDHNSVSSSENS 585
DB 758 EG-AQEEQVEPLDLSLPKQGBELLERIPLEQCLFCGERT-----797
QY 586 DEPLNLTPIKKEFSNSNLDNKSNNPVFGMNPFSAPKPLVTPLPQSAFPAPATMPVQTS 645
DB 798 ---LELVAKKEPQKOSCVTD--SEPVVAVIPPSANPINIAIPVTAQLPTIVALDQNS 852
QY 646 IGLRFPYGLQMSLPLHMAVTPYTGAAFTADMOQRKQKQKQKQKQKQKQKQKQKQKQK 705
DB 853 VPCFLALANKQITLIPVAYTYSTTSP-AVQEPPLKTIQPNQODEQODTSSGVSINV 911
QY 706 DDMTDSDSLRSKIKTKTESGMVACDLCDTQKSSSLRHXYETGKRPKPHOCQCKKAF 765
DB 912 EDQNDSDSTPRKGRKRTKENGMTACDLCDKTIQKSSSLRHXYETGKRPKPHOCQCKKAF 971
QY 766 KHKHHLIHSRLHSGEKPYQCDKCKRFSHSGSYQHMNHRYSYCKREAEEBAERBAR 825
DB 972 KHKHHLIHSRLHSGEKPYQCDKCKRFSHSGSSQHMNHRYSICKREAEEBDETEGE-- 1029
QY 826 EKGHGAPTELLMRAVLOSITP-QGYSDBEERESNPR--DGSEKKEHEGE-----CY 877
DB 1030 ---EAGP-EILSNHVGARASPOG--DSDERESTLTREDEDESEKEEBEDKEMELOEE 1083
QY 878 GKLRARDQDEEESEENKSMPTDEETIDEETGDSMDDESED-----GKM 928
DB 1084 KECEKPGQDEEEEBEVEBEVE-EAENBGEAKTEGLMDQDAEASASLGQKVES 1142
QY 929 ETKSDEEDN 938

```

DB 1143 SEQVSEKXTN 1152

RESULT 7

US-10-232-561-12
Sequence 12, Application US/10232561

Publication No. US20030119772A1

GENERAL INFORMATION:

APPLICANT: Genetia, Thomas

TITLE OF INVENTION: Methods and compositions useful for

FILE REFERENCE: CHOP 00-99

CURRENT FILING DATE: 2002-08-30

PRIOR APPLICATION NUMBER: 60/317,300

PRIOR FILING DATE: 2001-09-05

NUMBER OF SEQ ID NOS: 12

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 12

LENGTH: 1117

TYPE: PRT

ORGANISM: Mouse

```

QY 2 LTOGAGNRKFKCTGCGAKFKYKHHLKEHLRIHSGEKPEYECNCKKRFSGSHSISK 61
DB 210 VTOSGCKNRKFKCTGCGAKFKYKHHLKEHLRIHSGEKPEYECNCKKRFSGSHSISK 269
QY 62 KCIIGLISVNGRMRNNIKTG--SSPNSVSSPTNSAITQIRKLENGKPLMSBOGLIKI 119
DB 270 KCIISLIPVNGRERTIKTSQCSSP-SLSASPSGSPRPQIRKIEV-KPL--QEOULSVQI 326
QY 120 KTEPLDFNDYKVLMTATHGFGSSGSPMNGGLATSPGLGHPASQSMOHVAVGM 172
DB 327 KTEPLVDY-EFKPIVVASGINSCTPLQNGVFTGGGFLQATSSPQGVAVLFTVLVSPI 381
QY 177 LGFPTMNSLSEVQKVLQI-VDNTVSRQKMDCKTEDISLKGYHMKDPCSQPEOGVTSP 235
DB 440 -----SINLSDIQVNLKAVDGNVIRQVLETNQ---ASLASKEQEAVSASPIQGG 432
QY 236 NIPRVLGVVSHNGATKSIIDYTELEKNEAKACLOSITTSRQISNKKERKLTLDLV 291
DB 491 -ISATSLPLVDQDGTTKIINYSLEBPSQLQVVPQNLKKNPVATNSCKSEKLPED 491
QY 266 TD-DKMIENH-SISPTFGCOFCESFPGPI-PLHQHERYLCMNEIRKAVIOPHENI 349
DB 550 SEKDSFEGGVNDSTCLLCCDC---PGDL-----NALPELKHVDECPA 537
QY 353 KAGVFNVDKALLLSVLESEKGLTSPIN-PYQDHSVLKAYYANMNEPNSDELKISAVGL 408
DB 595 LPAEAEKPESSVSSATGDNLSPOQPLKXNLISLKKAYALNAPSAEELSKINDSVNL 595
QY 413 POEFYKEMFEQKRYQYNSRSPLERTSKPLAPNSNPTTKOSLIPRPVKPMDSITSPS 456
DB 655 PLDVKKMKFEKMOAQOIR--GQSDPSPGSGVNIPTKTBQPPAGNBPQEDST 654
QY 473 IAEHLNSTVSCDPLRLTKSSHFTNIKAVDKLDRSMSTPPLNLSSTSKSHSSSTTP 516
DB 701 ANEPDSTVTNLOSPLKMTNS---PVLPGSTTNGSRSSPTSPSLNLSRRTQGYLYTA 688
QY 533 NSFSEELQAEPLDLSLPKOME-----PKGIATKTKATKATINLDHNSVSSSENS 576
DB 758 EG-AQEEQVEPLDLSLPKQGBELLERIPLEQCLFCGERT-----TSSPSPLN 729
QY 586 DEPLNLTPIKKEFSNSNLDNKSNNPVFGMNPFSAPKPLVTPLPQSAFPAPATMPVQTS 631
DB 798 ---LELVAKKEPQKOSCVTD--SEPVVAVIPPSANPINIAIPVTAQLPTIVALDQNS 787
QY 646 IGLRFPYGLQMSLPLHMAVTPYTGAAFTADMOQRKQKQKQKQKQKQKQKQKQKQKQK 689
DB 853 VPCFLALANKQITLIPVAYTYSTTSP-AVQEPPLKTIQPNQODEQODTSSGVSINV

```

```

Db      | 788 ACPFTVAIADQNSVCLRALANKQTIIIPVAVAY---SATVSAVDEPPVKYIQPNG 844
Qy      | 690 FCGDLLDGAQDYVSGJDDMTDSCLSRKIKKTESGMYACDLCDKTFPKSSSLAHKXE 749
Db      | 845 NQDERDQTSSEGSIVYEDQNDSDSTPKKTKRTKTEGMYACDLCDKTFPKSSSLAHKXE 904
Qy      | 750 HFGKRPQOICCKAFKFKHHLIHSRLHSGEKPYQCDKCGKRFSGSGYSQHMNRYSY 809
Db      | 905 HGKGRPECCGICRKAFFKFKHHLIHSRLHSGEKPYQCDKCGKRFSGSGYSQHMNRYSY 964
Qy      | 810 CREAREPAAERAEAREKHLGPTL--LMNRAVLOSITPGYSDSESESMR--DGS 865
Db      | 965 CKRGADRDAMEQE---DAGEVLPVYLATEHVGRASPQADSDEREESTLREDEDS 1019
Qy      | 866 EKEHEKEGE---GYKLRRRGDDEEESESESNKSMDTDPETIRDEEFTGDH--- 917
Db      | 1020 EKEEEDDEMELEQSGKEENQSGEHEEVEVEVEVEVEVEVEVEVEVEVEVEVEVE 1079
Qy      | 918 -----SMDSEEDGKMETKSDHEE 936
Db      | 1080 VEVGAQAQAGSLERQKASEMESESESEQ 1108

```

```

RESULT 8
US-09-864-761-34862
; Sequence 34862, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/224,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203

```

```

; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34862
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010130.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 6.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6.8
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.7
; OTHER INFORMATION: EST_HUMAN HIT: BF245591.1, EVALUATE 1.00e-53
; OTHER INFORMATION: SWISSPROT HIT: P37275, EVALUATE 2.00e-33
US-09-864-761-34862

```

```

Query Match      20.1%; Score 1005; DB 9; Length 190;
Best Local Similarity 97.4%; Pred. No. 1.6e-51;
Matches 185; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy      | 753 KRPHOCQICCKAFKFKHHLIHSRLHSGEKPYQCDKCGKRFSGSGYSQHMNRYSYCKR 812
Db      | 1 KRPHOCQICCKAFKFKHHLIHSRLHSGEKPYQCDKCGKRFSGSGYSQHMNRYSYCKR 60
Qy      | 813 EAERERAEAREAREKHLGPTLNMNRAVLOSITPGYSDSESESMRDPGSEKEHEKE 872
Db      | 61 EAERERAEAREAREKHLGPTLNMNRAVLOSITPGYSDSESESMRDPGSEKEHEKE 120
Qy      | 873 GEGYCKLRRGDDEEESESESNKSMDTDPETIRDEEFTGDHSMDDSEDKMETKS 932
Db      | 121 GEGYCKLRGDDEEESESESNKSMDTDPETIRDEEFTGDHSMDDSEDKMETKS 180
Qy      | 933 DHEEDNMDG 942
Db      | 181 DHEEDNMDG 190

```

```

RESULT 9
US-09-864-761-36988
; Sequence 36988, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664

```



```

PRIORITY FILLING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 36988
LENGTH: 190
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC010130.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
OTHER INFORMATION: EST HUMAN HIT: BF245591.1, EVALUE 1.00e-53
OTHER INFORMATION: SWISSPROT HIT: P37275, EVALUE 2.00e-33
US-09-864-761-36988

Query Match      20.1%; Score 1005; DB 9; Length 190;
Best Local Similarity 97.4%; Pred. No. 1.6e-51;
Matches 185; Conservative 2; Mismatches 3; Indels 0; Gaps 0

QY 753 KRPHOCQCCKAFKKHHLIHSRLHSGEKPQCCKGCRFFSHSSSYSGHMHRYSYCKR 812
Db 1 KRPHOCQCCKAFKKHHLIHSRLHSGEKPQCCKGCRFFSHSSSYSGHMHRYSYCKR 60
QY EAEREAEAREAREKCHLGPTELLMNRAYLOSITPOGYSDSERESMPDGESEKEHEKE 872
Db 61 EAEREAEAREAREKCHLEPTELLNRRAYLOSIITOGYSDSERESMPDGESEKEHEKE 120
QY GEEGYGKLRRRDGDDEEEEEESENKSMDTPETIRDEELETGDSMDSSDGDMETKS 932
Db 121 GEDEYGLRGDRGDDEFEFEFESESNSKMDTPETIRDEELETGDSMDSSDGDMETKS 180
QY 933 DHEEDNMEDG 942
Db 181 DHEEDNMEDG 190

RESULT 10
US-10-087-340-83
Sequence 83, Application US/10097340
Publication No. US20030087250A1
GENERAL INFORMATION:
APPLICANT: John MONAHAN
APPLICANT: Manjula GANNAVARAPU
APPLICANT: Sebastian HOERSCH
APPLICANT: Shudhangt KAMWATKAR
```

APPLICANT: Steve G. KOVATS
 APPLICANT: Rachel E. MEYERS
 APPLICANT: Michael MORRISSEY
 APPLICANT: Peter OLANDT
 APPLICANT: Ami SEN
 APPLICANT: Peter VEIBY
 APPLICANT: Gordon B. MILLS
 APPLICANT: Robert C. BAST, Jr.
 APPLICANT: Karen LU
 APPLICANT: Rosemarie SCHMANDT
 APPLICANT: Xumei ZHAO
 APPLICANT: Karen GLATY
 TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification, Title Of Invention: Assessment, Prevention, and Therapy of Ovarian Cancer
 FILE REFERENCE: K01-030
 CURRENT FILING DATE: 2002-03-14
 PRIOR APPLICATION NUMBER: 60/276,025.
 PRIOR FILING DATE: 2001-03-14
 PRIOR APPLICATION NUMBER: 60/325,149
 PRIOR FILING DATE: 2001-09-26
 PRIOR APPLICATION NUMBER: 60/276,026
 PRIOR FILING DATE: 2001-03-14
 PRIOR APPLICATION NUMBER: 60/324,967
 PRIOR FILING DATE: 2001/09/26
 PRIOR APPLICATION NUMBER: 60/311,732
 PRIOR FILING DATE: 2001-08-10
 PRIOR APPLICATION NUMBER: 60/325,102
 PRIOR FILING DATE: 2001-09-26
 PRIOR APPLICATION NUMBER: 60/323,580
 PRIOR FILING DATE: 2001-09-19
 NUMBER OF SEQ ID NOS: 363
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 83
 LENGTH: 1042
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-097-340-83

Query Match 7.1%; Score 356.5; DB 15; Length 1042;
 Best Local Similarity 19.5%; Pred. No. 1.4e-12;
 Matches 205; Conservative 144; Mismatches 402; Indels 301; Gaps 38;

QY 9 RKEKTECGAFAKKYKHLKEHLRIHSGEKPECPNCKKRFSSHSGSYSHSKKCTGLIS 68
 Db 101 REYKCDQCPKAFNKSMLIRHOMSHDSCGHYECENCAVFPDPNLTQHIRSOHV----- 155
 QY 69 VNGMKRNKITGSSPNVSSTPTNSAITQLRNKLENGPL-----SMSQOTGLKIKT 121
 Db 156 -----GARAHACPECGKTPATSTSGLKOHGHHSVYPTICEVCHKSYTQFSNLCRHR 208
 QY 122 EPLD-----FN DYKVLMAATHGF-SGSSPFNNGIGATSP LGVHPSAQS PMOH 167
 Db 209 MHADCRQIKKCKDCQGFSTTSLANKARRFCEGKNHFAAGF----- 250
 QY 168 LGVGEAPRLGFPPTNSNSEVQKVLQIVDNTVSROKMDCKTEBISKLGYHMDPCSQP 227
 Db 251 FGGQISLP--GTPAMDK-----TSVWNNSHANPGADYFG-----ANR 286
 QY 228 EEQGVTSNPV-----PVGL-----PVSHNGATKSLIDTYLLEKNNAKACLOSLT 273
 Db 287 HPAGLITPTPAAGFSFSPGLRPPSGLYHRPPLIPASSPVKGL--SSTEQTNKSQS----- 338
 QY 274 TDSRQISNITKEKELRLIDLVTDDKMIENHSISTPFSCQCFCKSGFPPIPLHOHERYIC 333
 Db 339 -----PLMTHPIQLP 348
 QY 334 KMEIEKAVTIQPHENIVPNKAGVFNKALLLSVLSSEKGLTSPINPYKDHMSVYLKAYYA 393
 Db 349 ATQIILKA--LSKHPSVGDNK-----PVELQPPERSSEERPREK 384
 QY 394 MMPEPNSDELLKISIANGLPQEPFYKEMFEQKQVYQYSNRSRPSLERTSKPLAPNSNPFTTK 453

Db 385 ISDSESSDLDVSTPSSGDLFTT-----SSDLESIDSEKKEFKENGMKFK 432

Qy 454 DSLPRSVKMD-----SITSPSTAE---LHNSTSCDPLRLTKSHFTNIRKAVD 502

Db 433 DKVSPDLQNASINNKEYSNHSITFSPSLSEQTAAGVAANDSITALASIAEKYFGSTGLVG 492

Qy 503 KLDSRSNTSPPLNLSTSSKNSHSSSYTPNFSSEBELQAEPLD-----SLPQOM 553

Db 493 LQDKKVGMLPFP-SMFPPLFPFPFQSNYP--FPDRDLRLPLKMEPOSPEGVKLQJGS 549

Qy 554 REPKGIATKKTKATYSINLDHNSVSSSENSDEPLNLTFLIKFEESNNLIDNKSNNPVF 613

Db 550 SESFPDLTKKQKDEKPLTPVPSKPEVTPATSDQCLDLSMGRSPASGKLTPEAKHNVF 609

Qy 614 G-----NMFPKAPLPTLPPOSPAPPLATMPV-----QTSIP-----GLAPPY 652

Db 610 GKKGKSINVSRRPASQSL-----QHARPTPEFMDPIRVREKRKLTDLPLALEKXYLRPS 663

Qy 653 PGL---DOMSFLPHNAYTYPTGAATFADMOQRKRYQKQFQSGDLDDGADQVMSGLDMT 709

Db 664 PGFLFHPQMSAIEENNA-----EKLESISA-LKPEASSELQVPSMF 703

Qy 710 D---SDSCLSRKIKKTESGMVACDLCDTFOKSSSLRLHKYEHYTKRPHQCOLCKAKFK 766

Db 704 NFRAPNPLPENLRLRKGE-RYTCYCKGKIFPRSANLRLRLTHGEOQPYRKCYCDRSPS 762

Qy 767 HKHLIEHSR-LHSEKPYQCDKCKRPSHSGSYQHN-NRYSYCKRBAEREAEAREBA 824

Db 763 ISSNLQRHVRNIHNKEKPFKCHLDCRCFGQOQTNLDRHLKKHENGNMGSATSSPHSELES 822

Qy 825 -----REKCHIGP--TELLNRAVYQS-----ITPGYS--DSE 854

Db 823 TGAALDDKEDAYFTEIRNFIGNSHGSQSPNVEERRMNGSHKDEKALVTQNSDLDLDE 882

Qy 855 ERESMPRGESEKEKEHKEGEBGYKL---RRDGEDEEBEESSENNKMDTDPETIRDEE 912

Db 883 EVEDBEVLDDDEDDIDITGKTGKEVYTNLHGNEDDYEEISALBMSCKTSPVYKKEE 942

Qy 913 -ETGDHSM-----DSSEDKMETKSDHEED 937

Db 943 YKSGLSALDHIRHFTDSLKMRKMEQNSYSEAE 974

RESULT 11

US-10-097-340-79

Sequence 79, Application US/10097340

Publication No. US20030087250A1

GENERAL INFORMATION:

APPLICANT: John MONAHAN

APPLICANT: Manjula GANNAVARAPU

APPLICANT: Sebastian HOERSCH

APPLICANT: Shubhangi KAMATKAR

APPLICANT: Steve G. KOVATS

APPLICANT: Rachel E. MEYERS

APPLICANT: Michael MORRISSEY

APPLICANT: Peter OLANDT

APPLICANT: Ami SEN

APPLICANT: Peter VEIBY

APPLICANT: Gordon B. MILLS

APPLICANT: Robert C. BAST, JR.

APPLICANT: Karen LU

APPLICANT: Rosemarie SCHMANDT

APPLICANT: Xumel ZHAO

APPLICANT: Karen GLATT

TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification

TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer

FILE REFERENCE: MRI-030

CURRENT APPLICATION NUMBER: US/10/097,340

CURRENT FILING DATE: 2002-03-14

PRIOR APPLICATION NUMBER: 60/276,025

PRIOR FILING DATE: 2001-03-14

PRIOR APPLICATION NUMBER: 60/325,149

PRIOR FILING DATE: 2001-09-26

```

1. PRIOR APPLICATION NUMBER: 60/276,026
2. PRIOR FILING DATE: 2001-03-14
3. PRIOR APPLICATION NUMBER: 60/324,967
4. PRIOR FILING DATE: 2001/09/26
5. PRIOR APPLICATION NUMBER: 60/311,732
6. PRIOR FILING DATE: 2001-08-10
7. PRIOR APPLICATION NUMBER: 60/325,102
8. PRIOR FILING DATE: 2001-09-26
9. PRIOR APPLICATION NUMBER: 60/323,560
10. PRIOR FILING DATE: 2001-09-19
11. NUMBER OF SEQ ID NOS: 363
12. SOFTWARE: FastSeq for Windows Version 4.0
13. SEQ ID NO 79
14. LENGTH: 1051
15. TYPE: PRF
16. ORGANISM: Homo sapiens
17. US-10-097-340-79

Query Match          7.1%; Score 355.5; DB 15; Length 1051;
Best Local Similarity 19.4%; Pred. No. 1.6e-12;
Matches 199; Conservative 140; Mismatches 355; Indels 293; Gaps 36

QY      9  RKFKCTCGCAFKYKHHLKEHLRIHSEKPEPCNCKRFSHSGSYSHSKKCTGLIS 68
DB      101 REYKCDCCPRAFMKSNLIRHQMSHDGSKHYEGECNAKVFDPDNLQRHINSQV----- 155

QY      69  VNGMRNNITGSSPNSVSSPFINSAITOLANKLENGKPL-----SMSQTGLIKXT 121
DB      156 -----GAAAHACPECGKTFATPSTGKONKHGHSYKPFICEVCHKSXYTQFSMLCRHKR 208

QY      122 EPLD-----FNDYKVLMAHNGF--SGSSPFMNGGLATSPGLGVHPSAQSPPHQH 167
DB      209 MHACRTQIKCKDCGQMFSTTSSLNKRRPFCEGKNHPAAGGF----- 250

QY      168 LGVGEAPRLGFPFMNSNLSSEVKYLOIVNTVSRQKMDCKTEDISKLGYNMKDPCSQP 227
DB      251 FGGQISLP--GTPMDK-----TSMVMNSHANPGLADVFG-----ANR 286

QY      228 EEOGVTSNPPI-----PVGL---PVSHNGATKSIIDYLEKVNFAKACLOSLT 273
DB      287 HPALUTFPPTAPGFSFVSVGLPFGSLYHRPPLIPASSPVKGL--SSTEGTKNSQS----- 338

QY      274 TDSKROISNKKELKRLTLDLVTDMMIENHSISTPSCQCKSKSPFPIPLHQERLTC 333
DB      339 -----PLMTHPQILP 348

QY      334 KMEIEIKAVLOPHENIIVNKAGVEVDNKKALLSSVLSSEKGLTSPINPKDMYSLKAYYA 393
DB      349 ATQDILKA--LSKHSVGDNK-----PYELOPESSSEERPREK 384

QY      394 MNMEPNSDELKLSIAVGLPOEFVKEMFEORQVYQYSNSRPSLERTSKPLAPNSPTTK 453
DB      385 ISDSSESSLDLDVSTPGSDLETT-----SGSDLESIDIESDKFKENGKMKF 432

QY      454 DSLIPRSPVAKMD-----STSPSIAE---LHNSVTSQDPPRLKTSHPFNIAKYVD 502
DB      433 DKVSPLONLASINNKKEYSNHSISTSPSLHEQTAAGVANSIKAIASIAEKYFGSTGLVG 492

QY      503 KLDSRNTGSPPLNLSTSSKNSHSSSVTPSPFSSEBELQAEPLD-----SLPKQM 553
DB      493 LQDKKVALPYP--SMPLPFPFPAFSQSNYP--FPDRDRLSRJPLMKEPQSPQEVKKLQKGS 549

QY      554 REPKGIIATNKTKTATSTINLDHNSVSSSESDPEPLNTLFKEKFFSNMNLDNKSNPNVF 613
DB      550 SESEFIDLTKRKDKKPLPVPSKRPVTPATNSQDQPLDMSRSRASSTKLTERKKNHVF 609

QY      614 G-----MNPFAKPLYPPLPQSAFPPATPEPV-----QTSIP-----GLRPY 652
DB      610 GKKKGSNVESRPADGSL-----QHARPTPFMDPIRYAVEKRLTDLPLALKKCYLRPS 663

QY      653 PGLDQMSFLPMATYTPPGAAATPADMQQRKYQKKGQGGUULLDGAQDYMSGGLDMTP-- 710
DB      664 PG-----FLFPQQLDPORTMSAIEINMA--EKLESASA--LKDEASILLQSVPMENFR 715

```

```

Qy      711 -SDSCSRKKIKKTEGMAACDLCQTFQKSSLLRHKYETGKRPQCOICKAFGHK 769
Db      716 APPNALPENLTKGAE-RYTCRYCGKIPPRANLRLHRLTGTGQPRCKCDRSFIS 774

Qy      770 HLIHSR-LHSGEPYQCDGKGRPSHGYSQHMHRYSYCKRAEERAEAREAREKG 828
Db      775 NQHRNRNINHKKEPKCHLCYRCGQOOTNLDRH-----KKIENG 815

Qy      829 HLGPTELLNRAVLOSITPGYSDERESEMRDSESEKEHE--GEGYGLRRDQ 885
Db      816 NMSGT-----ATSSPHSELESGAILDDVEDAIFYEIRNFIGNSHGQSRRNV 864

Qy      886 DE-----EEBEESEENKSM---DTPETLRDEEETGDHMSDSEDEKMETYSKH 934
Db      865 EERMNNGSHKKEKALVPQNSDILLDEVEDEVLDEE---DEDYITGKTGKEPVTSNL 921

Qy      935 EEDNMD 941
Db      922 HEGNPD 928

```

```

RESULT 12
US-10-029-386-33060
; Sequence 33060, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33060
; LENGTH: 1204
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC003973.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.7
; OTHER INFORMATION: SWISSPROT HIT: O43345, EVALU0 0.00e+00
; US-10-029-386-33060

```

Query Match	7.0%;	Score 349.5;	DB 12;	Length 1204;
Best Local Similarity	20.5%;	Pred. No. 4.3e-12;		
Matches 192;	Conservative 126;	Mismatches 322;	Indels 295;	Gaps 36

Qy		6	AGNRKFKTECEGCKAFYKYXHLHKEHLRIHSIGCRPEPCPKRFS-----HSG--	52
Dd		416	AGEGRPYKCEEGCGAKFNMSSNLMEMHRITHGEKPYKCEEGCSFTSLTTHKAIIHTEGK	475
Qy		53	-----SYSHIISKKCICGLISVN-----GRMRN-----NIKTGSSP--	83
Dd		476	PYKCEEGCAKAYKMSSTLSYNHKKIHTVEKPYKCEEGCGAKFNQSAILIKHRIHTGEKPYKC	535
Qy		84	-----NSVSSPBTNSAIIQLRNKLBNKGRLMSEQGLLKITEPLDPN-----D	128
Dd		536	ECCGKTSKSOSTLTTHKAI-----HAGERPYKCEEGCKTFPIKVSTLTTHKAIIHAGEKP	588
Qy		129	YKVLMAATHGPS-----GSSEPFMNGILGA-----TSPLGVHP---SAOSPMDL	168
Dd		589	YKCEGCGAKAFSKFSILTGKVIHTGEKPYKCEEGCGAKFMSSNLMEMHRITHGEKPYKCE	648
Qy		169	GVGHEAPLPGFPPTMNSSNLSEVQKLVQIVD-----NTVRQRK-----DC-	207
Dd		649	ECGGSFSTSVLTTHKAIIHTEKPYKCEEGCGAKYKMSSTLSYNHKKIHTVEKPYKCEEGCK	708

[illegible]

```

13      RESULT 13
14      US-09-864-761-38036
15      ; Sequence 38036, Application US/09864761
16      ; Patent No. US20020048763A1
17      ; GENERAL INFORMATION:
18      ; APPLICANT: Penn, Sharon G.
19      ; APPLICANT: Rank, David R.
20      ; APPLICANT: Hanzel, David K.
21      ; APPLICANT: Chen, Wensheng
22      ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
23      ; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
24      ; FILE REFERENCE: Aeomica-X-1
25      ; CURRENT APPLICATION NUMBER: US/09/864,761
26      ; CURRENT FILING DATE: 2001-05-23
27      ; PRIOR APPLICATION NUMBER: US 60/180,312
28      ; PRIOR FILING DATE: 2000-02-04
29      ; PRIOR APPLICATION NUMBER: US 60/207,456
30      ; PRIOR FILING DATE: 2000-05-26
31      ; PRIOR APPLICATION NUMBER: US 09/632,366
32      ; PRIOR FILING DATE: 2000-08-03
33      ; PRIOR APPLICATION NUMBER: GB 24263.6
34      ; PRIOR FILING DATE: 2000-10-04
35      ; PRIOR APPLICATION NUMBER: US 60/236,359
36      ; PRIOR FILING DATE: 2000-09-27
37      ; PRIOR APPLICATION NUMBER: PCT/US01/00666
38      ; PRIOR FILING DATE: 2001-01-30
39      ; PRIOR APPLICATION NUMBER: PCT/US01/00667

```



```

; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1199
; LENGTH: 1357
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1199

```

```

Query Match      6.4%; Score 318; DB 12; Length 1357;
Best Local Similarity 18.7%; Pred. No. 3.5e-10;
Matches 163; Conservative 94; Mismatches 235; Indels 380; Gaps 28;

```

```

QY 7 GNRKCTECGKAFYKHLKEHLIHSGEKPECCPKRPSHSGSYSHISKKICGL 66
DB 693 GEKPYRCECGKAFWPSNLTRHKRIHTEKPYACECGQAFRRSSTLTNN----- 743
QY 67 ISVNGRMNNITGSSPNSVSSPTNSAITQLRNKLNGKPLSMSEQTGLKITEPLDF 126
DB 744 -----KRHTGERP----- 752
QY 127 NDYKYLMTATHGSPGSGSPFNGLGATSPGVHPSAQS PMQHLGVGMEAPLQPTMNSNL 186
DB 753 --YKCEBCKAFSVSSALI----- 769
QY 187 SEVQKVLQIVDNTVSRQKDKCTEDI SKLKGVMKDPSCQPEQGVTSNPVGLPVVS 246
DB 770 -----YHKR-----I 774
QY 247 HNGATKSIIDYTLKENVKAKACLOSLTDSRROISNIKEKRLTLDVTDKMIENNISI 306
DB 775 HTGER---PYTCECGKAFNCSTLKT-----HKII--HTG 805
QY 307 STPFCQCKESFPPIPLHOHER-----YLCNNBEIKAVLOPHENIVPNKAGVFD 359
DB 806 EKPYTCECGKRTFNCSSTYKAKHRIHTEKPYKC--EBCDKAFKWH----- 849
QY 360 NKALLLSVLSKGLTSPINPYKDHMSVLKAYYANMMEBNSDELIKISIAVGLPOEFYKE 419
DB 850 -----SLAKHKIHTGEKPYKC-----SDSKALAKSS----- 877
QY 420 WFEQKRYQ---YSNSRSPSLERTSKPLAPNSNPPTKOSLPRSPYKPMDSITSPSIAEL 476
DB 878 --EVQKVYSGDENIGIRVHKKETQGWLVRRNNENRTGLFOIRAAVRPV--RDPSSWQO 932
QY 477 HNSVTSQDPPLTLTK---SSHFTNIAKAVDK-----LDHS 507
DB 933 BGSILT--DPIQKEBPDLONHYDHQALBEDQRTNGVGLLTRDVTIERFSLBEWQCLDHA 990
QY 508 RSNTPSPNLSTSSKNSHSSSYT--PNSFSSELOAEPDL-----SLPKOM--R 554
DB 991 QONLVRDVLNRYMLVSLGIAVSKPDLITCLEQNKPEPNNIRNEMVTKHPDLPELIGIX 1050
QY 555 EPKGIATNKTAKAISINDHNSVSSSE-----NSDEPLNLTIFX----- 595
DB 1051 DSLQVYIPRRYKSGHDNLQVYTKCSMGECEVQKGCNEVNOCLSTQNKIPQTHKCVKV 1110
QY 596 -KEFSNNSNLDNKSNNPVFGMNPFAKPLVTLPPQSAFPATFMPVPVQTSIRGLRYPYG 654
DB 1111 RGFNSNR--HKTHT--GKHFKCKY----- 1135
QY 655 LDQMSFLPHMAYVYTPGATFADMQORRYQ---RKQFGDLDLDGAQDYNSGLDMDTD 710

```

```

DB 1136 -----GKSFQWVSQLOHQIHTRENSYQCE--ECGKPF-----N 1168
QY 711 SDSCISRKKIKTESGMVACDLCDKTFOKSSSLRHKYEHTGKRPPOCOICKAFKHH 770
DB 1169 CSSTLSKHKRIHTEKPYRCECGKAFWSSSTLTNRKRIHTEKPYTCECGQAFRRSST 1228
QY 771 LIEHSRLHSGEKPYQCDKCGRRFSGSGYSQH 802
DB 1229 LANKHRIHTEKPYTCECGKAFSLSSSLTVH 1260

```

Search completed: January 5, 2004, 16:37:15
Job time : 56 secs

THIS PAGE BLANK (USPTO)